

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2001, 11:35:01 ; Search time 20.31 seconds  
(without alignments)  
1095.470 Million cell updates/sec

Title: US-09-545-072A-2  
Perfect score: 1922  
Sequence: 1 MEAYKQWNRREYVQSGS.....EKIVLEGAQSRITYISGS 367

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_0601.\*  
1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1922	100.0	367	21 AAB19718	Arabidopsis thalia
2	249.5	13.0	336	21 AAB24227	Human vesicle asso
3	186.5	9.7	391	21 AAB19719	Yarrowia lipolytic
4	89.5	4.7	956	22 AAB79523	Corynebacterium gl
5	89.5	4.7	1257	17 AAR87628	Alpha-ketoglutaric
6	89.5	4.7	1257	19 AA411781	B. lactofermentum
7	89	4.6	3165	14 AAR38889	Sequence encoded b
8	88	4.6	1645	18 AAW13502	B. pertussis adeny
9	88	4.6	1705	10 AAP94365	Sequence of part o
10	88	4.6	1706	11 AAR08031	Adenyl cyclase fro
11	87.5	4.6	302	20 AAY05097	HIV Tat protein.

12	87.5	4.6	302	22 AAB87879	Synthetic fusion p
13	87	4.5	1522	10 AAP93357	Sequence of the ca
14	86.5	4.5	197	21 AAG52859	Arabidopsis thalia
15	86.5	4.5	288	21 AAG52858	Arabidopsis thalia
16	86.5	4.5	298	21 AAG52857	Arabidopsis thalia
17	86.5	4.5	1025	17 AA03185	Rice Xa21 disease
18	86.5	4.5	1025	20 AAW93597	O. longistaminata
19	86.5	4.5	1025	22 AAB67442	Amino acid sequenc
20	86.5	4.5	1026	20 AAY39279	Xa21 signal transd
21	85.5	4.4	2380	21 AAB18315	Plasmodium falci
22	85.5	4.4	2595	20 AAY39297	spnA a polyketide
23	85	4.4	424	20 AAY35407	Chlamydia pneumoni
24	84.5	4.4	1021	19 AAW76192	Actinoplanes sp. a
25	84.5	4.4	4150	21 AAY92707	S. antibioticus 8,
26	84	4.4	580	21 AAY82157	Pseudomonas putida
27	84	4.4	765	20 AAW89561	Recombinant heat-r
28	83.5	4.3	309	21 AAG38792	Arabidopsis thalia
29	83.5	4.3	460	21 AAG38791	Arabidopsis thalia
30	82.5	4.3	177	21 AAG24979	Arabidopsis thalia
31	82.5	4.3	197	21 AAG24978	Arabidopsis thalia
32	82.5	4.3	288	21 AAG24977	Arabidopsis thalia
33	82	4.3	220	21 AAG16938	Arabidopsis thalia
34	82	4.3	248	21 AAG16937	Arabidopsis thalia
35	82	4.3	252	21 AAG16936	Arabidopsis thalia
36	80.5	4.2	309	21 AAG17027	Arabidopsis thalia
37	80.5	4.2	460	21 AAG17026	Arabidopsis thalia
38	80.5	4.2	524	15 AAR56525	Protein kinase (NU
39	80.5	4.2	524	16 AAR76621	Saccharomyces cere
40	80.5	4.2	747	20 AAW89585	Human ATP-binding
41	80.5	4.2	747	21 AAB13355	Human ABC-Transpor
42	80.5	4.2	752	20 AAY14068	Arabidopsis thalia
43	80	4.2	400	21 AAG59923	Arabidopsis thalia
44	80	4.2	401	21 AAG59922	Arabidopsis thalia
45	80	4.2	730	16 AAR82881	Lupin exo-(1-4)bet

## ALIGNMENTS

RESULT 1  
AAB19718  
ID AAB19718 standard; Protein; 367 AA.  
XX  
AC AAB19718;  
XX  
DT 19-FEB-2001 (first entry)  
XX  
DE Arabidopsis thaliana SSE1 protein.  
XX  
KW SSE1; shrunken seed gene; storage reserve; storage protein;  
oil body; transgenic plant.  
XX  
OS Arabidopsis thaliana.  
XX  
FH Key  
FT Region Location/Qualifiers  
FT Region 38..53  
FT Region /note= "hydrophobic region"  
FT Region 109..124  
FT Region /note= "hydrophobic region"  
FT Region 134..180  
FT Region /note= "hydrophilic region"  
FT Region 201..227  
FT Region /note= "hydrophilic region"  
FT Region 238..257  
FT Region /note= "hydrophobic region"  
FT Region 337..355  
FT Region /note= "hydrophobic region"  
XX  
PN WO200061735-A1.  
XX  
PD 19-OCT-2000.  
XX  
PF 07-APR-2000; 2000WO-US09192.

```
XX PR 08-APR-1999; 99US-0128651.
XX XX (GEO ) GEN HOSPITAL CORP.
PA LIn Y;
XX WPI: 2000-679483/66.
XX N-PSDB; AAA88782.
XX Novel shrunken seed gene useful for producing transgenic plants having
PT altered production of food storage reserve material, intracellular
PT transport of storage protein and formation of protein or oil bodies
XX Claim 2; Page 58; 64pp; English.
XX The present sequence is that of Arabidopsis thaliana SSE1 (shrunken
CC seed) protein, as deduced from isolated cDNA (see AAA88782). SSE1
CC protein, when expressed in a cell of a plant, modifies or alters
CC the production of a food storage reserve material (e.g. protein,
CC lipid or carbohydrate storage reserve), facilitates the
CC intracellular transport of a storage protein, or facilitates a
CC formation of protein or oil bodies. The invention provides a
CC transgenic plant (or plant cell, plant tissue, plant organ or
CC plant component) which includes a recombinant SSE1 transgene that
CC modifies the production of food storage reserves, thereby
CC increasing nutritional value. An antisense construct is useful for
CC modifying desiccation tolerance.
XX Sequence 367 AA;
XX
Query Match 100.0%; Score 1922; DB 21; Length 367;
Best Local Similarity 100.0%; Pred. No. 6.2e-195; Mismatches 0; Indels 0; Gaps 0;
Matches 367; Conservative 0;
Qy 1 MEAYKQWVRNREYVQSGSFANGLTWLPKFSASEIGPEAVTAFIGFTTINEHIEN 60
Db 1 meaykqwwrrnreyvqsgsfangltwlpkfsaseigpeavtafigfttinehiien 60
Qy 61 APTPRGHVGGSGNDPSLSYPLLIALLKDLTVVVEVAEHFYGDKKWNYIILTEAMKAVIR 120
Db 61 aptprghvsgsgndpslsypylliallkdlctvvevaeahfygdkkwnyilteamkavir 120
Qy 121 LALFRNSGYKMLLOGETPNEEKDSNOSQNAGNSGRNLGPHGLGNQNHHPNLEGR 180
Db 121 lalfrnsykmllggetpnEEKdsnsqsnragnsgrnlghpqlgnqnhhpwnlegr 180
Qy 181 AMSALSSFGQNARTTSTTPGWSRRIOHQQAVIEPPMIKERRRTMSSELLTEKGVNGALFA 240
Db 181 amsalssfgqnarttsttpgwsrriohqqvavieppmikertrtmseiltekgvngalfa 240
Qy 241 IGEVLYITRPLTYLVLFIRKYGVRSWIPWALSISLSDVTLMGGLANSKWKGEKSKOVHFSGP 300
Db 241 lgevllyitrplylvlfirkygvrswipwalsisvdtlmgllanskwgkkskvhfsgp 300
Qy 301 EKDELRRRLKIWALYLMRDPFTTYTRQKLESSOKKLELIPLIGFLTEKIVELLEGAQSR 360
Db 301 ekdelrrrllkiwalylmrdpfttytrqklessckklelpligfltekiivellegaqsr 360
Qy 361 YTYISGS 367
Db 361 ytyisgs 367
RESULT 2
AAB24227
ID AAB24227 standard; protein; 336 AA.
XX AC AAB24227;
XX AC AAB24227;
XX DT 07-FEB-2001 (first entry)
XX
```

```
DE Human vesicle associated protein 6 SEQ ID NO:6.
XX Human; vesicle associated protein; VEAS; diagnosis; nootropic;
KW immunosuppressive; antibacterial; antihelminthic; protozoacide;
KW virucide; anti-inflammatory; cytostatic; gene therapy; akinesia; AIDS;
KW transport disorder; Bell's palsy; diabetes mellitus; catatonia;
KW infectious myositis; dementia; Grave's disease; autoimmune disorder;
KW inflammatory disorder; asthma; bronchitis; Reiter's syndrome; cancer;
KW systemic lupus erythematosus; rheumatoid arthritis; fungicide;
XX infection.
XX Homo sapiens.
OS WO200060082-A2.
XX PD 12-OCT-2000.
XX 06-APR-2000; 2000WO-US09353.
XX 07-APR-1999; 99US-0128193.
XX 20-JUL-1999; 99US-0144701.
XX (INCY-) INCYTE PHARM INC.
XX Lal P, Yue H, Hillman JL, Baughn MR, Tang YT, Lu DAM, Azimzai Y;
PI WPI: 2000-665006/64.
XX N-PSDB; AAC60990.
XX New vesicle associated proteins (VEAS) and polynucleotides encoding
PT VEAS for diagnosing, treating or preventing transport disorders (e.g.
PT Bell's palsy), autoimmune/inflammatory disorders (e.g. AIDS) or cancer
PT (e.g. Leukemia)
XX Claim 1; Page 84-85; 112pp; English.
XX AAC60995 to AAC61000 and AAC64367 to AAC64369 encode the human vesicle
XX associated proteins 1 to 19 (VEAS-1 to 19) given in AAB24222 to AAB24240.
XX Human VEAS proteins have nootropic, immunosuppressive, antibacterial,
CC fungicide, antihelminthic, protozoacide, virucide, anti-inflammatory and
CC cytostatic activities. The isolated polypeptides and polynucleotides are
CC useful for diagnosing, treating or preventing transport disorders (e.g.
CC akinesia, Bell's palsy, diabetes mellitus, catatonia, infectious
CC myositis, dementia, Grave's disease or AIDS), autoimmune/inflammatory
CC disorders (e.g. AIDS, asthma, bronchitis, Reiter's syndrome, systemic
CC lupus erythematosus, rheumatoid arthritis, or viral, fungal, bacterial,
CC protozoal or helminthic infections), or cancer (e.g. adenocarcinoma,
CC leukemia, melanoma, sarcoma, or gall bladder, heart, kidney, bone,
CC brain, breast, pancreas or skin cancers). The polypeptides may also be
CC used to immunise a host animal. The polypeptides and polynucleotides are
CC further useful for screening libraries of compounds or for drug
CC screening assays.
XX Sequence 336 AA;
XX
Query Match 13.0%; Score 249.5; DB 21; Length 336;
Best Local Similarity 25.4%; Pred. No. 8.2e-18; Mismatches 157; Indels 51; Gaps 12;
Matches 94; Conservative 68;
Qy 4 YKOWVRNREYVQSGSFANGLTWLPKFSASEIGPEAVTAFIGFTTINEHIENAPT 63
Db 11 yqeyvtrhpaaataqlatavrgfsyllagrfadshelselvysasnllvldngilrkelr 70
Qy 64 PRGHVGGSGNDPSLSYPL---IAILKDLTVVVEVAEHFYGD-KKWNVIILTEAMKAVI 119
Db 71 kklpv-----slsgqklltwlsvlecvevfmemgaakvwgdevgrwlviaqlakavl 123
Qy 120 RLALFRNSGYKMLLOGETPNEEKDSNOSQNAGNSGRNLGPHGLGNQNHHPNLEGR 179
Db 124 rmlll--lwfkaglg--tsppivplrdetqapddgd-----hspgnhe- 163
Qy 180 RAMSALSSFGQNARTTSTTPGWSRRIOHQQAVIEPPMIKERRRTM--SELLTEKGVNGA 237
```

Db 164 --qsyvkrsvrvrtlqntcp--slhsrhwa-----pqqregtqqdheelsatptplgl 215  
 Qy 238 LFAIGEVLITRPLIYVLFIRKYGVRSWIPWAISSLVSDTLGMLLANSKWWGKSKQVHF 297  
 Db 216 qetiaeflyiarplhlhlsiglgqrswwkpllagvvdvtsisllsdrkgltr----- 269  
 Qy 298 SGPEKDELRRKLWALYLMRDFFTKYTRQK-LESSQKLELIPLIGLFLTEKIVELLE 356  
 Db 270 ---errelrrtlllyllrsfydfeairllfillqlladhvpvgvltvtrplmdyldpt 326  
 Qy 357 AQRSTYVIG 366  
 Db 327 wkiyfyfswg 336  
 RESULT 3  
 ID AAB19719 standard; Protein; 391 AA.  
 AC AAB19719;  
 DT 19-FEB-2001 (first entry)  
 DE Yarrowia lipolytica Pexl6 protein.  
 KW SSE1; shrunken seed gene; storage reserve; storage protein;  
 KW oil body; transgenic plant.  
 OS Yarrowia lipolytica.  
 PN W0200061735-A1.  
 PD 19-OCT-2000.  
 PF 07-APR-2000; 2000WO-US09192.  
 PR 08-APR-1999; 99US-0128651.  
 PA (GEHO ) GEN HOSPITAL CORP.  
 PI Lin Y;  
 DR WPI; 2000-679483/66.  
 PT Novel shrunken seed gene useful for producing transgenic plants having  
 altered production of food storage reserve material, intracellular  
 transport of storage protein and formation of protein or oil bodies -  
 PS Disclosure; Page 59-60; 64pp; English.  
 CC The present sequence is that of Yarrowia lipolytica Pexl6 protein.  
 CC The invention includes a method for modifying or altering the  
 CC biosynthesis of a storage reserve in a transgenic plant (or its  
 CC component) by: introducing a transgene encoding an SSE polypeptide  
 CC linked to a promoter functional in plant cells; and regenerating a  
 CC transgenic plant from the transformed cells. The SSE polypeptide  
 CC is expressed in the cells of the transgenic plant, thereby  
 CC modifying or altering the seed storage reserve (lipid, storage  
 CC protein, or carbohydrate e.g. starch) of the plant. In preferred  
 CC embodiments, the expressed polypeptide is Pexl6.  
 XX Sequence 391 AA;

Query Match 9.7%; Score 186.5; DB 21; Length 391;  
 Best Local Similarity 20.7%; Pred. No. 4.9e-11;  
 Matches 86; Conservative 68; Mismatches 143; Indels 119; Gaps 13;  
 QY 1 MEAYKQWNRNRYVQSGFSGFANGLTWLLPEKFSASEIGPEAVTAPLGIFFITNEHIEN 60  
 Db 20 ldsydkflvrnaasigsistrtsvyvpgrfndveiatetlyavlrvlghlyhdtliar 79

QY 61 APTPRGHVSSGNDPSLSYP-----LLIAILKDLTVVEVAAEH 99  
 Db 80 a-----vaasnaavvypspbnrytdwifknrkgykysravtfvkgelvaemvakk 133  
 QY 100 FYGD-KKNYIILLTAMKAVIRLALFRNSGYKMLLOGGETPNEEK----- 143  
 Db 134 nggemaarkciligiegikagiriymlgstlyqplc---ttypdrevtgelleticrdeg 190  
 QY 144 ----DSNQSESONRAGSNRNLGPHGLGNQNHHPNWNLEGRAMSALSFSFGQNARTTTSST 199  
 Db 191 eldiekgldmpqknprrtgrtipea-----ptnvegylt----- 226  
 QY 200 PGWSRRHQQAIVIEPPMIKERRRTMSELTGKGVNGALFAIGEVLITRPLIYVLFIRK 259  
 Db 227 ----kvlrsevdvrdpynllsr-----ldnwg-----vaellsilrplyacilfr 268  
 QY 260 YGVR-----WIPWAISSLVSDTLGMLLANSKWWGKSKQVHSGP--- 300  
 Db 269 qhvnkvtvpastkskfpflnspwapwlglviealsrkmmgs---wllrrrq---sgktpt 322  
 QY 301 --EKDELRRKLWALYLMRDFFTKYTRQKLESSQKLELIPLIGLFLTEKIVELL 354  
 Db 323 aldgmevkgrtnllgwwlfrgefgyaytrpillysivarlekigpdlgfgalisdy 378  
 RESULT 4  
 ID AAB79523 standard; Protein; 956 AA.  
 AC AAB79523;  
 DT 30-APR-2001 (first entry)  
 DE Corynebacterium glutamicum SMP protein sequence SEQ ID NO:562.  
 KW Corynebacterium glutamicum: carbon metabolism and energy production;  
 KW SMP protein; sugar metabolism and oxidative phosphorylation protein;  
 KW fine chemical production; organic acid; proteinogenic amino acid;  
 KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;  
 KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;  
 KW carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;  
 KW diagnosis; Corynebacterium diphtheriae; evolutionary study.  
 OS Corynebacterium glutamicum.  
 PN WO200100844-A2.  
 PD 04-JAN-2001.  
 PF 23-JUN-2000; 2000WO-IB00943.  
 XX 25-JUN-1999; 99US-0141031.  
 PR 08-JUL-1999; 99DE-1031412.  
 PR 08-JUL-1999; 99DE-1031413.  
 PR 08-JUL-1999; 99DE-1031419.  
 PR 08-JUL-1999; 99DE-1031420.  
 PR 08-JUL-1999; 99DE-1031424.  
 PR 08-JUL-1999; 99DE-1031428.  
 PR 08-JUL-1999; 99DE-1031431.  
 PR 08-JUL-1999; 99DE-1031433.  
 PR 08-JUL-1999; 99DE-1031434.  
 PR 08-JUL-1999; 99DE-1031510.  
 PR 08-JUL-1999; 99DE-1031562.  
 PR 08-JUL-1999; 99DE-1031634.  
 PR 09-JUL-1999; 99DE-1032180.  
 PR 09-JUL-1999; 99DE-1032227.  
 PR 09-JUL-1999; 99DE-1032230.  
 PR 09-JUL-1999; 99US-0143208.  
 PR 14-JUL-1999; 99DE-1032924.  
 PR 14-JUL-1999; 99DE-1032973.  
 PR 14-JUL-1999; 99DE-1033005.  
 PR 27-AUG-1999; 99DE-1040765.  
 PR 31-AUG-1999; 99US-0151572.

PR 03-SEP-1999; 99DE-1042076.  
 PR 03-SEP-1999; 99DE-1042079.  
 PR 03-SEP-1999; 99DE-1042086.  
 PR 03-SEP-1999; 99DE-1042087.  
 PR 03-SEP-1999; 99DE-1042088.  
 PR 03-SEP-1999; 99DE-1042095.  
 PR 03-SEP-1999; 99DE-1042123.  
 PR 03-SEP-1999; 99DE-1042125.  
 XX  
 PA (BADI ) BASF AG.  
 XX  
 PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;  
 XX  
 DR WPI; 2001-061975/07.  
 DR N-PSDB; AAF71640.  
 XX  
 PT New isolated Corynebacterium glutamicum nucleic acid encoding a sugar  
 PT metabolism and oxidative phosphorylation protein for production or  
 PT modulation of production of fine chemicals e.g. amino acids,  
 PT carbohydrates or enzymes -  
 XX  
 PS Claim 20; Page 944-946; 1246pp; English.  
 XX  
 CC AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar  
 CC metabolism and oxidative phosphorylation (SMP) proteins given in  
 CC AAB79243 to AAB 79633 which are involved in carbon metabolism and  
 CC energy production. The C. glutamicum SMP gene can be used in vectors  
 CC (II) for expression in host cells and production or modulation of  
 CC production of fine chemicals, such as, an organic acid, a proteinogenic  
 CC or nonproteinogenic amino acid (preferred), a purine or pyrimidine base,  
 CC a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty  
 CC acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a  
 CC cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins  
 CC (III) encoded by them are used for diagnosing the presence or activity of  
 CC Corynebacterium diptheriae in a subject. (I), (II), (III) or host cells  
 CC containing them are used to map genomes of organisms related to  
 CC C. glutamicum, identify and localise C. glutamicum sequences of interest,  
 CC in evolutionary studies, in determining SMP protein regions required  
 CC for function, in modulating SMP protein activity, in modulating the  
 CC metabolism of sugars, and in modulating high-energy molecule production  
 CC in a cell (i.e. ATP, NADPH).  
 XX  
 SQ Sequence 956 AA;

Query Match 4.7%; Score 89.5; DB 22; Length 956;  
 Best Local Similarity 20.2%; Pred. No. 3.7; Indels 85; Gaps 13;  
 Matches 62; Conservative 37; Mismatches 123;  
 Qy 65 RGHVSSGNDPSLSYPLLIAILKDLTVVEVAAEHFYGDKKWNYIILTEAMKAVIRLALF 124  
 Db 480 rgh--neaddpsmtqpkmyellitgretvraqytedlllgrd----lsnedaavvr--df 531  
 Qy 125 RNSGYKMLLOGGETPNEEKDSNQSESONRAGNSGRNLGPHGLGNQ----- 169  
 Db 532 hd-----qmesvfnvkeggkkaeaqgitgsqkphgletnlsreellelqgafan 584  
 Qy 170 -----NHNH-----PWNLGRAMSALSSFGQNARTTTSSTPGWSR 204  
 Db 585 tpegfnyhrvpavakkrvsvtggidawg-ellafgslansgrlvraged-----sr 639  
 Qy 205 R--IQHQOAVIEPPMIKERRRTMSSELLTEKGVNGALFAIGEVLVITRPLIYVLFIRKYG 262  
 Db 640 rgtftqrhavaidpataeefnplhelaqskgnngkf-----lvynsaletayagmgfeygy 694  
 Qy 263 R-----SWIPWALSISVDTLGMGLL-----ANSKMWGEKSKOV-----HFSGPEKDE 304  
 Db 695 svgnedsivaeaqfgdfangaqtiideyvsgeakwgqtsklllllphgyegggpohss 754  
 Qy 305 LRRRKLII 311  
 Db 755 arierfl 761

RESULT 5  
 AAR87628  
 ID AAR87628 standard; Protein; 1257 AA.  
 XX  
 AC AAR87628;  
 XX  
 DT 02-AUG-1996 (first entry)  
 XX  
 DE Alpha-ketoglutaric acid dehydrogenase.  
 XX  
 KW L-glutamic acid; L-lysine; production; Coryneform bacteria;  
 KW alpha-ketoglutaric acid dehydrogenase; efficient.  
 XX  
 OS Brevibacterium lactofermentum.  
 XX  
 PN W09534672-A.  
 XX  
 PD 21-DEC-1995.  
 XX  
 PF 07-JUN-1995; 95WO-JP011131.  
 XX  
 PR 14-JUN-1994; 94JP-0131744.  
 XX  
 PA (AJIN ) AJINOMOTO CO INC.  
 XX  
 PI Abe C, Asakura Y, Kawahara Y, Kimura E, Kurahashi O;  
 PI Nakamatsu T, Tsujimoto N, Usuda Y;  
 XX  
 DR WPI; 1996-049699/05.  
 DR N-PSDB; AAT08998.  
 XX  
 PT Coryneform L-glutamic acid producing bacteria - useful in producing  
 PT L-glutamic acid and L-lysine  
 XX  
 PS Claim 4; Page 35-44; 62pp; Japanese.  
 XX  
 CC The present sequence is that of an alpha-ketoglutaric acid dehydrogenase  
 CC derived from L-glutamic acid producing Coryneform bacteria. L-glutamic  
 CC producing Coryneform bacteria are useful in producing L-glutamic acid.  
 CC Coryneform bacteria contg. the DNA and L-lysine producing ability lead  
 CC to the production of L-lysine.  
 XX  
 SQ Sequence 1257 AA;

Query Match 4.7%; Score 89.5; DB 17; Length 1257;  
 Best Local Similarity 20.2%; Pred. No. 5.7; Indels 85; Gaps 13;  
 Matches 62; Conservative 37; Mismatches 123;  
 Qy 65 RGHVSSGNDPSLSYPLLIAILKDLTVVEVAAEHFYGDKKWNYIILTEAMKAVIRLALF 124  
 Db 781 rgh--neaddpsmtqpkmyellitgretvraqytedlllgrd----lsnedaavvr--df 832  
 Qy 125 RNSGYKMLLOGGETPNEEKDSNQSESONRAGNSGRNLGPHGLGNQ----- 169  
 Db 833 hd-----qmesvfnvkeggkkaeaqgitgsqkphgletnlsreellelqgafan 885  
 Qy 170 -----NHNH-----PWNLGRAMSALSSFGQNARTTTSSTPGWSR 204  
 Db 886 tpegfnyhrvpavakkrvsvtggidawg-ellafgslansgrlvraged-----sr 940  
 Qy 205 R--IQHQOAVIEPPMIKERRRTMSSELLTEKGVNGALFAIGEVLVITRPLIYVLFIRKYG 262  
 Db 941 rgtftqrhavaidpataeefnplhelaqskgnngkf-----lvynsaletayagmgfeygy 995  
 Qy 263 R-----SWIPWALSISVDTLGMGLL-----ANSKMWGEKSKOV-----HFSGPEKDE 304  
 Db 996 svgnedsivaeaqfgdfangaqtiideyvsgeakwgqtsklllllphgyegggpohss 1055  
 Qy 305 LRRRKLII 311  
 Db 1056 arierfl 1062









16-JUN-1999 (first entry)  
HIV Tat protein.  
HIV; Tat protein; epitope; human immunodeficiency virus; immune response;  
viral multiplication reduction; chronic viraemia.  
Human immunodeficiency virus type 1.  
WO902185-A1.  
21-JAN-1999.  
10-JUL-1998; 98WO-0514332.  
11-JUL-1997; 97US-0893853.  
(THYM-) THYMON LLC.  
Goldstein G;  
WPI; 1999-120516/10.  
N-PSDB; AAX28268.  
Peptides containing epitopes of human immune deficiency virus Tat  
protein - used to induce antibodies reacting with HIV-1 Tat  
proteins, used to, e.g. reduce HIV-1 levels in both infected and  
uninfected patients  
Claim 20; Page 77-78; 140pp; English.

This sequence represents the HIV Tat protein, fragments of this sequence  
are claimed for use in the composition of the invention.  
The invention relates to a composition containing a peptide of formulae  
(I)-(IV): R1-VDP-y-LEP-R2 (I); R3-K-x-LGISGRKK-R4 (II);  
R5-RR-x'-zay'-S-R6 (III); R7-SQ-x''-HQ-y''-SLSKQP-R8 (IV); where:  
y = R, K, S, N or ornithine; R1, R3, R5 and R7 = hydrogen, lower alkyl or  
alkanoyl, or 1-5 (for R1 and R3) or 1-3 (for R5 and R7) amino acids (aa),  
optionally substituted by lower alkyl or alkanoyl; R2 = free hydroxy,  
amide or 1-4 additional aa, optionally substituted by lower alkyl or  
alkanoyl; x = G or A; R4 = hydroxy, amide or 1-5 additional aa, G or S;  
optionally substituted by amide; x' = A, P, S or Q; y' = D, N, G or S;  
z = P or H; a = Q or P; R6 = hydroxy or amide; x'' = N or T;  
y'' = A or V, and R8 = hydroxy, amide or 1-3 additional aa, optionally  
substituted by amide. The compositions, and synthetic molecules, viruses,  
or bacteria containing them, are used to induce Ab that react with over  
95, best 99, % of known human immunodeficiency virus (HIV)-1 Tat  
proteins, particularly to reduce HIV-1 levels in both infected and  
uninfected subjects (provided they are still immunocompetent). Antibodies  
(Ab) against the peptides reduce viral multiplication during initial  
infection and minimise the chronic viraemia that leads to acquired immune  
deficiency syndrome. Ab are used to reduce HIV-1 levels in infected  
subjects unable to mount an immune response (passive immunisation).  
Peptides (I)-(IV) represent four different epitopes in Tat protein, and a  
relatively small number of them will induce antibodies cross-reactive  
with most strains and variants of Tat. The content of the composition is  
easily adjusted to account for and new strains discovered. They can be  
administered in the early stages of, or before, infection and do not  
promote selection of escape mutants.

Sequence 302 AA;

Query Match 4.6%; Score 87.5; DB 20; Length 302;  
Best Local Similarity 24.1%; Pred. No. 1;  
Matches 45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;  
QY 64 PRGHVSSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123  
DB 91 pwehpgssgvdprl-----epwnhl-----gssgvdhrlep 121  
QY 124 FRNSGYKMLLOGGETPNEE-----KDSNQESQNRAGNSG---RNLLGPHGLGN-ON 170

DB 122 wkhpqsgdlqrtrtpqdsgrqrrrppqdsgrqrrrppqsgsgrqrrppqsgsgrqr 181  
QY 171 HHNPWNLEG-----RAMSALSSFGQGNARTTTSSTPGWRRRIHQHOAVIEPP 216  
DB 182 rrpqnsgrqrrrppqdsgrqrrrppqnsgrqrrrtpqdsgrqrrrahnqs----- 236  
QY 217 MIKERRR 223  
DB 237 gsrqrrr 243  
RESULT 12  
AAB87879  
ID AAB87879 standard; protein; 302 AA.  
XX  
AC AAB87879;  
XX  
DT 17-MAY-2001 (first entry)  
XX  
DE Synthetic fusion protein.  
XX  
KW Tat; HIV-1; human immunodeficiency virus; AIDS; viremia.  
XX  
OS Unidentified.  
XX  
PN US6193981-B1.  
XX  
PD 27-FEB-2001.  
XX  
PF 10-JUL-1998; 98US-0113921.  
XX  
PR 11-JUL-1997; 97US-0893853.  
XX  
PA (THYM-) THYMON LLC.  
XX  
PI Goldstein G;  
XX  
DR WPI; 2001-243400/25.  
XX  
PT New compositions comprising at least two variants of a Tat protein  
coupled to a carrier protein, useful for impairing multiplication of  
human immunodeficiency virus 1 and in eliciting anti-TAT antibodies  
PS Disclosure; Fig 2; 63pp; English.  
XX  
CC The present invention relates to a composition with at least two  
variants of Tat protein coupled to a carrier protein. The invention is  
useful for inhibiting the multiplication of HIV-1 virus in infected  
patients with viremia, symptomatic or asymptomatic and for  
attenuating HIV-1 multiplication during primary infection in  
previously uninfected subjects.  
XX  
SQ Sequence 302 AA;  
Query Match 4.6%; Score 87.5; DB 22; Length 302;  
Best Local Similarity 24.1%; Pred. No. 1;  
Matches 45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;  
QY 64 PRGHVSSGNDPSLSYPLLIALLKDLTVVEAAEHFYGDKKNYIILTEAMKAVIRLAL 123  
DB 91 pwehpgssgvdprl-----epwnhl-----gssgvdhrlep 121  
QY 124 FRNSGYKMLLOGGETPNEE-----KDSNQESQNRAGNSG---RNLLGPHGLGN-ON 170  
DB 122 wkhpqsgdlqrtrtpqdsgrqrrrppqdsgrqrrrppqsgsgrqrrppqsgsgrqr 181  
QY 171 HHNPWNLEG-----RAMSALSSFGQGNARTTTSSTPGWRRRIHQHOAVIEPP 216  
DB 182 rrpqnsgrqrrrppqdsgrqrrrppqnsgrqrrrtpqdsgrqrrrahnqs----- 236  
QY 217 MIKERRR 223

```

Db      237 gsrgrir 243
:::|:|:|
RESULT 13
AAP93357
ID AAP93357 standard; protein; 1522 AA.
XX AC
XX AC AAP93357;
XX DT
XX DT 22-MAR-1991 (first entry)
XX DE
XX DE Sequence of the catalytic domain of Bordetella pertussis
XX DE adenylate cyclase (AC).
XX KW
XX KW Vaccine; ss.
XX OS
XX OS Bordetella pertussis.
XX PN
XX PN FR2618453-A.
XX PD
XX PD 27-JAN-1989.
XX PF
XX PF 24-JUL-1987; 87FR-0010589.
XX PR
XX PR 24-JUL-1987; 87FR-0010589.
XX (INSP ) INST PASTEUR.
XX PA
XX PA Danchin A, Glaser P, Ullmann A;
XX PI
XX PI WPI; 1989-079098/11.
XX DR
XX DR N-PSDB; AAN90659.
XX
XX Cloning and expressing genes in multi-protein function system - by
XX PT transforming host cell with indicator gene producing protein which
XX PT interacts with expression prod. of the gene being cloned.
XX PS
XX PS Claim 14; Fig 2(a-f); 24pp; French.
XX CC
XX CC Sequences encoding at least a part of AC able to hybridise with a
XX CC gene expressing a protein with AC activity, and proteins with AC
XX CC activity able to form immune complexes with specific antibodies are
XX CC claimed. The isolated AC gene can be used to produce Bordetella
XX CC pertussis toxin.
XX SQ
XX SQ Sequence 1522 AA;

Query Match 4.5%; Score 87; DB 10; Length 1522;
Best Local Similarity 21.7%; Pred. No. 14;
Matches 70; Conservative 39; Mismatches 104; Indels 110; Gaps 17;

QY 25 LTWLLPKFSASIGPEAVTAF-----LGFTTINEHIENAPTGRGHV----- 68
Db 274 llwkiaagarsavgtearrqfygdgmigvitdf-elevnlnrrrhavagadvvqh 332
QY 69 GSSGNDPSLSYPLLIALLKDLTVVEVAEAEHYGDKKNYIILTEAMKAVIRLALFRNSG 128
Db 333 gteqnp---fp-----eadek---ifvsa-----tg 354
QY 129 YKMLQGG---ETPNEKDSNQESQNRRA-GNSGRNLGPHGIGNQNHHPNWNLEGRAMSA 184
Db 355 esqmlrqlkeyigqrggyvfyenraygvagskslfdgigaa-----pgvpsgr----- 406
QY 185 LSSFGQARTTTSPTGWSRRHQHQAIVLEPPMIKERRRTWSELLEKGVNGALFAIGEV 244
Db 407 -skfspdletypaspplrr-----pslgaverqdsydsldgvgsrfsisgev 454
QY 245 -----LITRPLIYLVFIRKYGRVSWIPWAISSVDTLGMGLANSKWMGEKSKQ 294
Db 455 sdmaaveaaeentrqviha-----garq-----ddaepgvsgsaahwgqralq 498

```

```

QY 295 VHFSGPEKDELRRRLKIWALYLM 317
Db 499 ----gaqavaaagr-llhaialm 516
:|:|:|
RESULT 14
AAG52859
ID AAG52859 standard; Protein; 197 AA.
XX AC
XX AC AAG52859;
XX DT
XX DT 18-OCT-2000 (first entry)
XX DE
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 67238.
XX KW
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS
XX OS Arabidopsis thaliana.
XX PN
XX PN EP1033405-A2.
XX PD
XX PD 06-SEP-2000.
XX PF
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126264.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 08-APR-1999; 99US-0128714.
XX PR 16-APR-1999; 99US-0129845.
XX PR 19-APR-1999; 99US-0130077.
XX PR 21-APR-1999; 99US-0130449.
XX PR 23-APR-1999; 99US-0130510.
XX PR 28-APR-1999; 99US-0130891.
XX PR 30-APR-1999; 99US-0131449.
XX PR 30-APR-1999; 99US-0132048.
XX PR 04-MAY-1999; 99US-0132407.
XX PR 05-MAY-1999; 99US-0132484.
XX PR 06-MAY-1999; 99US-0132485.
XX PR 06-MAY-1999; 99US-0132486.
XX PR 06-MAY-1999; 99US-0132487.
XX PR 07-MAY-1999; 99US-0132863.
XX PR 11-MAY-1999; 99US-0134256.
XX PR 14-MAY-1999; 99US-0134218.
XX PR 14-MAY-1999; 99US-0134219.
XX PR 14-MAY-1999; 99US-0134221.
XX PR 14-MAY-1999; 99US-0134370.
XX PR 18-MAY-1999; 99US-0134768.
XX PR 19-MAY-1999; 99US-0134941.
XX PR 20-MAY-1999; 99US-0135124.
XX PR 21-MAY-1999; 99US-0135353.
XX PR 24-MAY-1999; 99US-0135629.
XX PR 25-MAY-1999; 99US-0136021.
XX PR 27-MAY-1999; 99US-0136392.
XX PR 28-MAY-1999; 99US-0136782.
XX PR 01-JUN-1999; 99US-0137222.
XX PR 03-JUN-1999; 99US-0137528.
XX PR 04-JUN-1999; 99US-0137502.
XX PR 07-JUN-1999; 99US-0137724.
XX PR 08-JUN-1999; 99US-0138094.
XX PR 10-JUN-1999; 99US-0138540.
XX PR 14-JUN-1999; 99US-0138847.
XX PR 16-JUN-1999; 99US-0139119.
XX PR 16-JUN-1999; 99US-0139452.
XX PR 17-JUN-1999; 99US-0139453.
XX PR 17-JUN-1999; 99US-0139492.

```

PR 18-JUN-1999;	99US-0139454.	PR 16-AUG-1999;	99US-0149368.
PR 18-JUN-1999;	99US-0139455.	PR 17-AUG-1999;	99US-0149175.
PR 18-JUN-1999;	99US-0139456.	PR 18-AUG-1999;	99US-0149426.
PR 18-JUN-1999;	99US-0139457.	PR 20-AUG-1999;	99US-0149722.
PR 18-JUN-1999;	99US-0139458.	PR 20-AUG-1999;	99US-0149723.
PR 18-JUN-1999;	99US-0139459.	PR 20-AUG-1999;	99US-0149729.
PR 18-JUN-1999;	99US-0139460.	PR 23-AUG-1999;	99US-0149902.
PR 18-JUN-1999;	99US-0139461.	PR 23-AUG-1999;	99US-0149930.
PR 18-JUN-1999;	99US-0139462.	PR 25-AUG-1999;	99US-0150566.
PR 18-JUN-1999;	99US-0139463.	PR 26-AUG-1999;	99US-0150884.
PR 18-JUN-1999;	99US-0139464.	PR 27-AUG-1999;	99US-0151065.
PR 18-JUN-1999;	99US-0139750.	PR 27-AUG-1999;	99US-0151066.
PR 21-JUN-1999;	99US-0139763.	PR 27-AUG-1999;	99US-0151080.
PR 21-JUN-1999;	99US-0139817.	PR 30-AUG-1999;	99US-0151303.
PR 22-JUN-1999;	99US-0139899.	PR 31-AUG-1999;	99US-0151438.
PR 23-JUN-1999;	99US-0140353.	PR 01-SEP-1999;	99US-0151930.
PR 23-JUN-1999;	99US-0140354.	PR 07-SEP-1999;	99US-0152363.
PR 24-JUN-1999;	99US-0140695.	PR 10-SEP-1999;	99US-0153070.
PR 24-JUN-1999;	99US-0140823.	PR 13-SEP-1999;	99US-0153758.
PR 29-JUN-1999;	99US-0140991.	PR 15-SEP-1999;	99US-0154018.
PR 30-JUN-1999;	99US-0141287.	PR 16-SEP-1999;	99US-0154039.
PR 01-JUL-1999;	99US-0141842.	PR 20-SEP-1999;	99US-0154779.
PR 01-JUL-1999;	99US-0142154.	PR 22-SEP-1999;	99US-0155139.
PR 02-JUL-1999;	99US-0142055.	PR 23-SEP-1999;	99US-0155486.
PR 06-JUL-1999;	99US-0142390.	PR 24-SEP-1999;	99US-0155659.
PR 08-JUL-1999;	99US-0142803.	PR 28-SEP-1999;	99US-0156458.
PR 09-JUL-1999;	99US-0142920.	PR 29-SEP-1999;	99US-0156596.
PR 12-JUL-1999;	99US-0142977.	PR 04-OCT-1999;	99US-0157117.
PR 13-JUL-1999;	99US-0143342.	PR 05-OCT-1999;	99US-0157753.
PR 14-JUL-1999;	99US-0143624.	PR 06-OCT-1999;	99US-0157865.
PR 15-JUL-1999;	99US-0144005.	PR 07-OCT-1999;	99US-0158029.
PR 16-JUL-1999;	99US-0144085.	PR 08-OCT-1999;	99US-0158232.
PR 16-JUL-1999;	99US-0144086.	PR 12-OCT-1999;	99US-0158369.
PR 19-JUL-1999;	99US-0144325.	PR 13-OCT-1999;	99US-0159293.
PR 19-JUL-1999;	99US-0144331.	PR 13-OCT-1999;	99US-0159294.
PR 19-JUL-1999;	99US-0144332.	PR 13-OCT-1999;	99US-0159295.
PR 19-JUL-1999;	99US-0144333.	PR 14-OCT-1999;	99US-0159329.
PR 19-JUL-1999;	99US-0144334.	PR 14-OCT-1999;	99US-0159330.
PR 19-JUL-1999;	99US-0144335.	PR 14-OCT-1999;	99US-0159331.
PR 20-JUL-1999;	99US-0144352.	PR 14-OCT-1999;	99US-0159637.
PR 20-JUL-1999;	99US-0144632.	PR 14-OCT-1999;	99US-0159638.
PR 20-JUL-1999;	99US-0144884.	PR 18-OCT-1999;	99US-0159584.
PR 21-JUL-1999;	99US-0144814.	PR 21-OCT-1999;	99US-0160741.
PR 21-JUL-1999;	99US-0145086.	PR 21-OCT-1999;	99US-0160767.
PR 21-JUL-1999;	99US-0145088.	PR 21-OCT-1999;	99US-0160768.
PR 22-JUL-1999;	99US-0145085.	PR 21-OCT-1999;	99US-0160770.
PR 22-JUL-1999;	99US-0145087.	PR 21-OCT-1999;	99US-0160814.
PR 22-JUL-1999;	99US-0145089.	PR 21-OCT-1999;	99US-0160815.
PR 22-JUL-1999;	99US-0145192.	PR 22-OCT-1999;	99US-0160980.
PR 23-JUL-1999;	99US-0145145.	PR 22-OCT-1999;	99US-0160981.
PR 23-JUL-1999;	99US-0145218.	PR 22-OCT-1999;	99US-0160989.
PR 23-JUL-1999;	99US-0145224.	PR 25-OCT-1999;	99US-0161404.
PR 26-JUL-1999;	99US-0145276.	PR 25-OCT-1999;	99US-0161405.
PR 27-JUL-1999;	99US-0145913.	PR 25-OCT-1999;	99US-0161406.
PR 27-JUL-1999;	99US-0145918.	PR 26-OCT-1999;	99US-0161359.
PR 27-JUL-1999;	99US-0145919.	PR 26-OCT-1999;	99US-0161360.
PR 28-JUL-1999;	99US-0145951.	PR 26-OCT-1999;	99US-0161361.
PR 02-AUG-1999;	99US-0146386.	PR 26-OCT-1999;	99US-0161362.
PR 02-AUG-1999;	99US-0146388.	PR 28-OCT-1999;	99US-0161920.
PR 02-AUG-1999;	99US-0146389.	PR 28-OCT-1999;	99US-0161922.
PR 03-AUG-1999;	99US-0147038.	PR 28-OCT-1999;	99US-0161992.
PR 04-AUG-1999;	99US-0147204.	PR 28-OCT-1999;	99US-0161993.
PR 04-AUG-1999;	99US-0147302.	PR 29-OCT-1999;	99US-0162142.
PR 05-AUG-1999;	99US-0147192.		
PR 05-AUG-1999;	99US-0147260.		
PR 06-AUG-1999;	99US-0147303.		
PR 06-AUG-1999;	99US-0147416.		
PR 09-AUG-1999;	99US-0147493.		
PR 09-AUG-1999;	99US-0147935.		
PR 10-AUG-1999;	99US-0148319.		
PR 11-AUG-1999;	99US-0148319.		
PR 12-AUG-1999;	99US-0148341.		
PR 13-AUG-1999;	99US-0148565.		
PR 13-AUG-1999;	99US-0148684.		

Query Match 4.5%; Score 86.5; DB 21; Length 197;  
Best Local Similarity 31.1%; Pred. No. 0.66; 1; Gaps 1;  
Matches 19; Conservative 16; Mismatches 25; Indels 1;

QY 55 EHIIENAPTRGHVSSGNDPSLSYPLLIATLKDLFTVVEAAEHFYGDKNWYILTEA 114  
:t:ll:: : :l : : :ll: : :l : :lll : :l : :llll :ll  
Db 85 qhliedstlvkvigidgsvklfndygsi-kvdedlsdlnqkiggdkkwglslltet 143  
QY 115 M 115  
Db 144 i 144

RESULT 15  
AAG52858  
ID AAG52858 standard; Protein; 288 AA.  
XX AC AAG52858;  
XX DT 18-OCT-2000 (first entry)  
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 67237.  
XX KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX OS Arabidopsis thaliana.  
XX PN EP1033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-0301439.  
XX PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 08-APR-1999; 99US-0128234.  
PR 16-APR-1999; 99US-0128714.  
PR 19-APR-1999; 99US-0129845.  
PR 21-APR-1999; 99US-0130077.  
PR 23-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134376.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 14-JUN-1999; 99US-0138847.  
PR 16-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 15-JUL-1999; 99US-0143624.  
PR 16-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 19-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.









GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2001, 00:49:05 ; Search time 119.83 Seconds  
(without alignments)  
7770.826 Million cell updates/sec

Title: US-09-545-072a-1  
Perfect score: 1483  
Sequence: 1 attgcaacaggagagaaaa.....attaaaaaaaaaaaaaaa 1483

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues  
Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_0601.\*  
1: /SID88/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID88/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID88/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID88/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID88/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID88/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID88/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID88/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID88/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID88/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID88/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID88/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID88/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID88/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID88/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID88/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID88/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID88/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID88/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID88/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID88/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SID88/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1483	100.0	1483	21	Arabidopsis thalia
2	85.4	5.8	936	22	Oligonucleotide D1
3	85.4	5.8	936	22	Oligonucleotide D1
4	85.4	5.8	936	22	Oligonucleotide D1
5	85.4	5.8	936	22	Oligonucleotide D1
6	85.4	5.8	936	22	Oligonucleotide D1
7	85.4	5.8	936	22	Oligonucleotide D1
8	85.4	5.8	936	22	Oligonucleotide D1
9	83.8	5.7	936	22	Oligonucleotide D1
10	83.8	5.7	936	22	Oligonucleotide D1
11	83.8	5.7	936	22	Oligonucleotide D1

12	83.8	5.7	936	22	AAF58262	Oligonucleotide D2
13	83.8	5.7	936	22	AAF58255	Oligonucleotide D1
14	51.2	3.5	244	22	AAF58238	Oligonucleotide D1
15	45.6	3.1	244	22	AAF58238	Oligonucleotide D1
16	42	2.8	2070	21	AAC76167	Human ORFX ORF1722
17	41.6	2.8	2123	18	AAT59701	PTH-like peptide D
18	40.4	2.7	603	21	AAA16288	Human colon cancer
19	40.2	2.7	199	22	AAF17640	Human breast cancer
20	40.2	2.7	887	21	AAC59297	Human secreted pro
21	39.6	2.7	1361	21	AAA26399	Human secreted pro
22	39.6	2.7	1547	21	AAC59424	Human secreted pro
23	39.6	2.7	1738	21	AAC99089	Human pancreatic c
24	39.4	2.7	441	20	AAC09474	PL35-NT3 construct
25	38.4	2.7	1528	18	AAT90733	Human consensus b1
26	39.4	2.7	4120	20	AAZ09473	Human RNA helicase
27	39	2.6	396	22	AAF94827	Human ovarian can
28	39	2.6	721	21	AAC73912	Human secreted pro
29	39	2.6	2833	21	AAC77810	Human cancer assoc
30	39	2.6	3021	21	AAC75886	Human ORFX ORF1441
31	38.8	2.6	1495	21	AAZ7985	Corn CCR4 transcri
32	38.8	2.6	2364	21	AAZ33323	Human secreted pro
33	38.4	2.6	225	21	AAC98734	Human colon cancer
34	38.4	2.6	302	21	AAC98700	Human colon cancer
35	38.4	2.6	775	19	AAV44861	Clone BV66.1 codin
36	38.4	2.6	1671	21	AAC59315	Human secreted pro
37	38.2	2.6	2670	21	AAF18202	Lung cancer associ
38	38	2.6	3047	21	AAZ07586	Drosophila piwi ge
39	38	2.6	7286	20	AAZ57224	WO 9923223 Seq ID
40	38	2.6	7938	20	AAZ57235	WO 9923223 Seq ID
41	37.8	2.5	818	20	AAZ16631	Human gene express
42	37.8	2.5	818	20	AAZ16629	Human gene express
43	37.8	2.5	1087	21	AAC32912	Arabidopsis thalia
44	37.8	2.5	1437	21	AAC59567	Human secreted pro
45	37.6	2.5	248	21	AAC95297	Cat flea head and

## ALIGNMENTS

RESULT	1
AAA88782	
ID	AAA88782 standard; cDNA; 1483 BP.
AC	AAA88782;
XX	
DF	19-FEB-2001 (first entry)
XX	
DE	Arabidopsis thaliana SSE1 cDNA.
XX	
KW	SSE1; shrunken seed gene; storage reserve; storage protein;
KW	oil body; transgenic plant; ss.
XX	
OS	Arabidopsis thaliana.
XX	
FH	Key
FT	Location/Qualifiers
FT	CDS
FT	122..1225
XX	/*tag= a
PN	WO200061735-A1.
XX	
PD	19-OCT-2000.
XX	
PF	07-APR-2000; 2000WO-US09192.
XX	
PR	08-APR-1999; 99US-0128651.
XX	
PA	(GEHO ) GEN HOSPITAL CORP.
XX	
PI	Lin Y;
XX	
DR	WPI; 2000-679483/66.
DR	P-ESDB; AAB19718.
XX	

PT Novel shrunken seed gene useful for producing transgenic plants having  
 PT altered production of food storage reserve material, intracellular  
 PT transport of storage protein and formation of protein or oil bodies -  
 XX  
 XX Claim 8; Page 57; 64pp; English.

XX The present sequence is that of Arabidopsis thaliana SSEL (shrunken  
 CC seed) cDNA, which was isolated from a seedling cDNA library using  
 CC a partial genomic clone as probe. SSEL encodes a protein (see  
 CC AAB19718) that, when expressed in a cell of a plant, modifies or  
 CC alters the production of a food storage reserve material (e.g.  
 CC protein, lipid or carbohydrate storage reserve), facilitates the  
 CC intracellular transport of a storage protein, or facilitates the  
 CC formation of protein or oil bodies. The invention provides a  
 CC transgenic plant (or plant cell, plant tissue, plant organ or  
 CC plant component) which includes a recombinant SSEL transgene that  
 CC modifies the production of food storage reserves, thereby  
 CC increasing nutritional value. An antisense construct is useful for  
 CC modifying desiccation tolerance.

SQ Sequence 1483 BP; 461 A; 268 C; 347 G; 407 T; 0 other;

Query Match 100.0%; Score 1483; DB 21; Length 1483;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1483; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 attgcaaccaggagagaagaataatcaagattgatttaacgtgaatgaattttgtg 60  
 Db 1 attgcaaccaggagagaagaataatcaagattgatttaacgtgaatgaattttgtg 60  
 QY 61 ttcccaaaattctctgagaaatagcaagtttcagttttgtttctctctatctgaagctc 120  
 Db 61 ttcccaaaattctctgagaaatagcaagtttcagttttgtttctctctatctgaagctc 120  
 QY 121 aatggaagcttaatacaatgggtttggagaaatagagatgtatgatactctttggatc 180  
 Db 121 aatggaagcttaatacaatgggtttggagaaatagagatgtatgatactctttggatc 180  
 QY 181 ctttgcaacggattgacacgtctctctctgagaaagtttctctcagagattggacc 240  
 Db 181 ctttgcaacggattgacacgtctctctctgagaaagtttctctcagagattggacc 240  
 QY 241 agaagcagtaacggcttttttggcattattcaacagataataatgaacataattgaaaa 300  
 Db 241 agaagcagtaacggcttttttggcattattcaacagataataatgaacataattgaaaa 300  
 QY 301 tgcctcaaacctcgtggccattgttgatcttcgggaatgatccatccctttctattcc 360  
 Db 301 tgcctcaaacctcgtggccattgttgatcttcgggaatgatccatccctttctattcc 360  
 QY 361 actactatcgccatctcaagatttgaaactgttggaagtgccagctgaacactt 420  
 Db 361 actactatcgccatctcaagatttgaaactgttggaagtgccagctgaacactt 420  
 QY 421 ctatggagacaaaatggaaactacattattctcactgaagctatgaaagctgtcattag 480  
 Db 421 ctatggagacaaaatggaaactacattattctcactgaagctatgaaagctgtcattag 480  
 QY 481 gttagccttttcgggaatagtggtatagatgcttctcagagagggaacacactaa 540  
 Db 481 gttagccttttcgggaatagtggtatagatgcttctcagagagggaacacactaa 540  
 QY 541 tgaggagaagaattcttaaccaatccagctgcgaataatagagctggtaattcgggtagaaa 600  
 Db 541 tgaggagaagaattcttaaccaatccagctgcgaataatagagctggtaattcgggtagaaa 600  
 QY 601 tctcggccctcatggttttggaacaaatcatcatatcatgaacttggaagcagc 660  
 Db 601 tctcggccctcatggttttggaacaaatcatcatatcatgaacttggaagcagc 660  
 QY 661 ggcagatgctgttaagttcatttggatgagaatgcaagaacaaacacatctctacc 720  
 Db 661 ggcagatgctgttaagttcatttggatgagaatgcaagaacaaacacatctctacc 720

Db 661 ggcagatgctgttaagttcatttggatgagaatgcaagaacaaacacatctctacc 720  
 QY 721 cggttgctcgaagaattcaaacatcagcagcagtttatagagctccaatgatcaagga 780  
 Db 721 cggttgctcgaagaattcaaacatcagcagcagtttatagagctccaatgatcaagga 780  
 QY 781 gaggcgaagaacgatgtccgagctactactactgagaagggtgttaattgagcgtgtttgc 840  
 Db 781 gaggcgaagaacgatgtccgagctactactactgagaagggtgttaattgagcgtgtttgc 840  
 QY 841 gatgggtgaggtttttacataacagacacgcctcatttaagttttttcatcagaataa 900  
 Db 841 gatgggtgaggtttttacataacagacacgcctcatttaagttttttcatcagaataa 900  
 QY 901 tggagtcgcatcttgatctctctgggtctatctcttcttgagacacactggggtgag 960  
 Db 901 tggagtcgcatcttgatctctctgggtctatctcttcttgagacacactggggtgag 960  
 QY 961 tcttctgcaaatcgaagtgtggtgggagagaagcaagcaagtcctatttctcaggacc 1020  
 Db 961 tcttctgcaaatcgaagtgtggtgggagagaagcaagcaagtcctatttctcaggacc 1020  
 QY 1021 tgaagaagatgagctgagagacacaaactgatatgggcttgcctctacctcatgagatcc 1080  
 Db 1021 tgaagaagatgagctgagagacacaaactgatatgggcttgcctctacctcatgagatcc 1080  
 QY 1081 attctccaccaagtacacagcaggaagctggaagctctcaagaagctggaactaat 1140  
 Db 1081 attctccaccaagtacacagcaggaagctggaagctctcaagaagctggaactaat 1140  
 QY 1141 tccattgctgatttctccacagagaagattgtgagcttttgaggagctcagtcacg 1200  
 Db 1141 tccattgctgatttctccacagagaagattgtgagcttttgaggagctcagtcacg 1200  
 QY 1201 gtacattacatctcggatcgtgaggttaagcgttttacttatggtttatgacacgg 1260  
 Db 1201 gtacattacatctcggatcgtgaggttaagcgttttacttatggtttatgacacgg 1260  
 QY 1261 aagaattatgcttgggaatgcttttttagatcatcaaaagctctcagatcttctt 1320  
 Db 1261 aagaattatgcttgggaatgcttttttagatcatcaaaagctctcagatcttctt 1320  
 QY 1321 agggaaatgggttcaggctttttgttagaaattgttttattcaacaggtagaacaataa 1380  
 Db 1321 agggaaatgggttcaggctttttgttagaaattgttttattcaacaggtagaacaataa 1380  
 QY 1381 caatagacagatgtatctgaagagataagcttctctatgtctaaagaaatggacgatac 1440  
 Db 1381 caatagacagatgtatctgaagagataagcttctctatgtctaaagaaatggacgatac 1440  
 QY 1441 gaataaaacaagcatcattaaagattaaaaaataaaaaa 1483  
 Db 1441 gaataaaacaagcatcattaaagattaaaaaataaaaaa 1483

RESULT 2  
 AAF58252/c  
 ID AAF58252 standard; DNA; 936 BP.  
 XX  
 AC AAF58252;  
 XX  
 XX 24-APR-2001 (first entry)  
 XX  
 DE Oligonucleotide D1835.  
 XX  
 KW Electron-transfer group; ETM; mismatch; genotyping;  
 KW gene expression; ss.  
 XX  
 OS Synthetic.  
 XX  
 PN WO200107665-A2.  
 XX  
 PD 01-FEB-2001.

```

XX PF      26-JUL-2000; 2000MO-US20476.
XX PR      26-JUL-1999;   99US-0145695.
XX PR      17-MAR-2000; 2000US-0190259.
XX PA      (CLIN-) CLINICAL MICRO SENSORS INC.
XX PI      Umek RX;
XX DR      WPI; 2001-159728/16.
XX PT      Nucleic acids containing electron-transfer group, useful as labels in
XX PT      hybridization assays, e.g. for genotyping, allowing repeat analyses on
XX PT      a single surface -
XX PS      Example 6; Page 127; 159pp; English.
XX CC      The present invention relates to a composition comprising two nucleic
CC CC      acids each containing an electron-transfer group (ETM) having
CC CC      different redox potentials. The invention is used for electronic
CC CC      detection of nucleic acids, especially of substitutions (mismatches)
CC CC      and single-nucleotide polymorphisms, e.g. for genotyping,
CC CC      monitoring gene expression.
XX SQ      Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match          5.8%; Score 85.4; DB 22; Length 936;
Best Local Similarity 0.7%; Pred. No. 2.4e-13;
Matches 5; Conservative 440; Mismatches 306; Indels 0; Gaps

QY 733 aagaattcaacatcagcaagcaggittatagacctccaatgatcatgaagagcgaaagaac 792
Db 762 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 703
QY 793 gatgtccgagctacttactgagaagggtttaatggagcgttgtttgcgatttgtagagt 852
Db 702 WWWWWWWWWWWWWGWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 643
QY 853 tctttacaacagcagccgctatttcagcttctttcacagaaaatatggagtcgatac 912
Db 642 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 583
QY 913 ttgattctctgggctatatcgctttctgtggacacactggggatgggtctcttcgcaa 972
Db 582 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 533
QY 973 ttcgaagtgtgggagagaagacaagcaagtcattctcaagacctgaaagatga 1032
Db 522 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 463
QY 1033 gctgaggacgaaactgatatggcgatttacctcatgagagatccattcttcacca 1092
Db 462 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 403
QY 1093 gtacacaagcagaagctggaagctctcaaaagaagctggaactaatccattgatcgg 1152
Db 402 WWWWWTTAAGCWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 343
QY 1153 attctccaagagaagattgtgagcttttggaggagctcagtcagtcggtacacttacat 1212
Db 342 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 283
QY 1213 atcgggatcgtgaggttaagcgttttacttatggtttatatgcaacggaagaataattgcc 1272
Db 282 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 223
QY 1273 attgttggaaatgctttttatagatcatcaaaagcctccacagatttcttaagggaatttt 1332
Db 222 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 163
QY 1333 caggcttttgttagaaattgtgtttatttgcaacaggtagagaacataaacattagacagat 1392

```

```
QY 913 ttgattccttggtggtatcgcttctgtgagacactggggtggtcttcttgcaa 972
Db 582 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 523
QY 973 ttcgaagtgtgggagagagagcaagcaagtcattctcaggacctgaaagatga 1032
Db 522 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 463
QY 1033 gctgaggagacgaaactgatggtcattgtacccatgagagatccattcttccaa 1092
Db 462 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 403
QY 1093 gtacacagcagcaagctgaaagctctcaaaagagctggaactaatccattgatcg 1152
Db 402 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 343
QY 1153 attcctcacagagaagattgtgagcttttggaggagctcagtcacgtacattacat 1212
Db 342 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 283
QY 1213 atcgggagtcgagggttaagcgttttacttatgttttatgcaacggaagaattgcc 1272
Db 282 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 223
QY 1273 attgttggaatgcttttttagatcatcaaaagctcctacagattcttagggaattgtt 1332
Db 1332 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 163
QY 1333 caggctttttagaattgtttatttgcacaggtagagacaataaccatagacagat 1392
Db 162 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 103
QY 1393 gtatctgaagagataagcttctctatgtctaaagaaatgaccgatacagaataaacaag 1452
Db 102 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 43
QY 1453 catcattaaagattaaaaaataaaaaa 1483
Db 42 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 43
```

## RESULT 4

```
AAF58257/c
ID AAF58257 standard; DNA; 936 BP.
XX AC AAF58257;
XX AC 24-APR-2001 (first entry)
XX DT Oligonucleotide D1954.
XX DE Electron-transfer group; ETM; mismatch; genotyping;
XX KW gene expression; sg.
XX OS Synthetic.
XX OS WO200107665-A2.
XX PN 01-FEB-2001.
XX PD 26-JUL-2000; 2000WO-US20476.
XX PF 26-JUL-1999; 99US-0145695.
XX PR 17-MAR-2000; 2000US-0190259.
XX PR (CLIN-) CLINICAL MICRO SENSORS INC.
XX PA Umek RM;
XX PI WPI; 2001-159728/16.
XX DR Nucleic acids containing electron-transfer group, useful as labels in
XX PT
```

```
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX Example 6; Page 127; 159pp; English.
XX The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;
```

```
Query Match 5.8%; Score 85.4; DB 22; Length 936;
Best Local Similarity 0.7%; Pred. No. 2.4e-13;
Matches 5; Conservative 440; Mismatches 306; Indels 0; Gaps 0;
```

```
QY 733 aagaattcaacatcagcagcagttatagagcctcaatgagagagagcgagaac 792
Db 762 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 703
QY 793 gatgtccagctactactgagagaggtgttaagagcggtttgttgcatgttgaggt 852
Db 702 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 643
QY 853 tctttacataacgagcgcctcatttacgttcttttcacagaaatattgagtcgcatc 912
Db 642 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 583
QY 913 ttggattccttggtggtatcgcttctgtgacacatcggttggtcttcttgcaaa 972
Db 582 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 523
QY 973 ttcgaagtgtgggagagagagcaagcaagtcattctcagacacctgaaagatga 1032
Db 522 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 463
QY 1033 gctgaggagacgaaactgatggtgcatgttacctcatgagagatccattcttccacaa 1092
Db 462 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 403
QY 1093 gtacacagcagcaagctgaaagctctcaaaagagctggaactaatccattgatcg 1152
Db 402 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 343
QY 1153 attcctcacagagaagattgtgagcttttggaggagctcagtcacgtacactacat 1212
Db 342 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 283
QY 1213 atcgggagtcgagggttaagcgttttacttatgttttatgcaacggaagaattgcc 1272
Db 282 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 223
QY 1273 attgttggaatgcttttttagatcatcaaaagctcctacagattcttagggaattgtt 1332
Db 222 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 163
QY 1333 caggctttttagaattgtttatttgcacaggtagagacaataaccatagacagat 1392
Db 162 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 103
QY 1393 gtatctgaagagataagcttctctatgtctaaagaaatgaccgatacagaataaacaag 1452
Db 102 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 43
QY 1453 catcattaaagattaaaaaataaaaaa 1483
Db 42 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 12
```

RESULT 5









[illegible]



[illegible]

WWWWWWW T7 gg





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2001, 00:18:40 ; Search time 96.91 Seconds  
(without alignments)  
2835.013 Million cell updates/sec

Title: US-09-545-072A-1  
Perfect score: 1483  
Sequence: 1 attgcaaccaggaagagaaa.....attaaaaaaaaaaaaaaaaa 1483

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 317530 seqs, 92630169 residues

Total number of hits satisfying chosen parameters: 635060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PCRUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	38.8	2.6	7218	1	US-08-232-463-14
2	38	2.6	7286	4	US-09-331-581-3
3	38	2.6	7938	4	US-09-331-581-14
4	37.2	2.5	1117	4	US-09-247-373B-33
5	37.2	2.5	3871	2	US-08-599-455B-3
6	36.8	2.5	2837	2	US-08-993-228-11
7	36.4	2.5	2547	3	US-08-262-220-7
8	36.4	2.5	2547	3	US-08-471-733-7
9	36.4	2.5	2547	3	US-08-468-878-7
10	36.4	2.5	2547	3	US-08-750-494-7
11	36	2.4	940	2	US-08-471-717-1
12	36	2.4	2433	1	US-08-136-743B-40
13	36	2.4	2663	1	US-08-136-743B-3
14	35.6	2.4	8920	2	US-08-446-855A-1
15	35.6	2.4	8920	4	US-09-150-741-1
16	35.4	2.4	240	1	US-08-628-417-6
17	35.4	2.4	4137	3	US-09-221-235-1
18	35.4	2.4	4137	4	US-09-221-928-1
19	35.4	2.4	4137	4	US-09-221-527-1
20	35.4	2.4	4137	4	US-09-221-236-1
21	35.4	2.4	4137	4	US-09-221-416-1
22	35.4	2.4	4137	4	US-09-221-245-1
23	35.4	2.4	4137	4	US-09-163-115-1
24	35.4	2.4	4137	4	US-09-221-528-1
25	35.4	2.4	4137	4	US-09-593-553-1
26	35.4	2.4	4137	4	US-09-221-237-1
27	35	2.4	1493	1	US-08-340-820-24

28	35	2.4	1493	1	US-08-593-535-24
29	35	2.4	1582	3	US-08-545-196B-10
30	35	2.4	1582	3	US-08-545-198B-12
31	35	2.4	5173	1	US-08-242-677-1
32	35	2.4	6671	1	US-08-280-443-1
33	35	2.4	6671	1	US-08-457-459-1
34	35	2.4	6671	1	US-08-555-678-1
35	35	2.4	6671	5	PCN-US95-02275-1
36	35	2.4	72928	3	US-09-009-913-1
37	34.8	2.3	1882	4	US-09-370-253-1
38	34.6	2.3	1296	1	US-08-385-191A-1
39	34.6	2.3	1441	4	US-08-821-994-63
40	34.4	2.3	746	4	US-09-013-810-1
41	34.4	2.3	1525	4	US-09-157-603-4
42	34.4	2.3	1525	4	US-09-587-436-4
43	34.2	2.3	144	1	US-08-702-344-26
44	34.2	2.3	2422	1	US-07-867-106-5
45	34.2	2.3	3138	1	US-07-867-106-4

ALIGNMENTS

RESULT 1  
US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232.463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION NUMBER: US/07/935.313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMM  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pTZgpt-F1s  
US-08-232-463-14





0.

**RESULT** 5

sequence 3, Application US/08599455B2  
Patent No 5072621

GENERAL INFORMATION:

APPLICANT: **Tepper, Robert I.**

**TITLE OF INVENTION: METHODS OF**

NUMBER OF SEQUENCES: 44

; ADDRESS: Fish &amp; Ric

; CITY: Boston

COUNTRY: U

COMPUTER READABLE I

COMPUTER: IBM Compatible

OPERATING SYSTEM: WINDOWS95  
SOFTWARE: FastSEO for Windows

APPLICATION NUMBER: IIS/08/599.455B

FILED DATE: 22-JAN-1996  
PRIOR APPLICATION DATA:

REPLICATION NUMBER: 08/383,153  
FILING DATE: 28-DEC-1995

APPLICATION NUMBER: 08/370,142  
FILING DATE: 11-DEC-1995

APPLICATION NUMBER: 08/569,485  
FILING DATE: 08-DEC-1995

APPLICATION NUMBER: 08/566,622  
FILING DATE: 04-DEC-1995

APPLICATION NUMBER: 08/562,663  
FILING DATE: 27-NOV-1995

ATTORNEY/AGENT INFORMATION:

REGISTRATION NUMBER: 35,283

**TELECOMMUNICATION INFORMATION:**

TELEFAX: 617-542-8906

INFORMATION FOR SEQ ID NO: 3:

LENGTH: 3871 base pairs

**STRANDEDNESS:** doub

**MOLECULE TYPE:** cDNA

•

Query Match 2.5%; Score 36.8; DB 2; Length 2837;  
Best Local Similarity 47.0%;  
Matches 110; Conservative 1; Mismatches 123; Indels 0;  
Pred. No. 0.38;

Query Match 2.5%; Score 36.8; DB 2; Length 2837;  
Best Local Similarity 47.0%; Pred. NO. 0.38;  
Matches 110; Conservative 1; Mismatches 123; Indels 0; Gaps 0;

QY 1250 atatgaacggaagaatattgcatgttgggaatgcttttttagatcatcaaaaggctcct 1309  
 Db 2594 ATGGGCTGTGAATAATTAATTTGGAGTTTGTATAAGAACATTTTAATTTGACTTAATATTTT 2653  
 QY 1310 acagattcttaggaatggtttcaggctttttagaattgtttattgcaacaggt 1369  
 Db 2654 AAGATGTGAATCCAGATCACTTGTCTGTTTAAATGGAATGCTTTTACAGGAGCTGT 2713  
 QY 1370 agagaacataaccatgacagatgtatctggaagagataagctctctatgtctaaagaaa 1429  
 Db 2714 AACATAMTTAAAAATATGAATGATTTATGTATGTAATATGCTTCTTTACATAAAAAATAAAA 2773  
 QY 1430 tggaccatgaataaaacacacatcattaaagattataaaaaa 1483  
 Db 2774 TGCAACACTGTAAAAA 2827

RESULT 7  
 ; Sequence 7, Application US/08262220  
 ; Patent No. 6054296

GENERAL INFORMATION:  
 APPLICANT: BERGSTROM SVEN  
 APPLICANT: BARBOUR ALAN G.  
 TITLE OF INVENTION: NEW 66 KDA ANTIGEN FROM BORRELIA  
 NUMBER OF SEQUENCES: 14  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: BROWDY AND NEIMARK  
 STREET: 419 SEVENTH STREET, N.W.  
 CITY: WASHINGTON  
 COUNTRY: USA  
 ZIP: 20004

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/262,220  
 FILING DATE: 20-JUN-1994  
 CLASSIFICATION: 514  
 ATTORNEY/AGENT INFORMATION:  
 NAME: COOPER, IVER P.  
 REGISTRATION NUMBER: 28,005  
 REFERENCE/DOCKET NUMBER: BERGSTROM=3  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 202-628-5197  
 TELEFAX: 202-737-3528  
 TELEX: 248633

INFORMATION FOR SEQ ID NO: 7:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2547 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 ORIGINAL SOURCE:  
 ORGANISM: Borrelia garinii  
 STRAIN: Ip90  
 IMMEDIATE SOURCE:  
 CLONE: PJB-101  
 NAME/KEY: CDS  
 LOCATION: 380..2245  
 FEATURE:  
 NAME/KEY: sig\_peptide  
 LOCATION: 380..442  
 FEATURE:  
 NAME/KEY: mat\_peptide  
 LOCATION: 443..2242  
 US-08-262-220-7

Query Match 2.5%; Score 36.4; DB 3; Length 2547;  
 Best Local Similarity 47.4%; Pred. NO. 0.47;  
 Matches 109; Conservative 0; Mismatches 121; Indels 0; Gaps 0;  
 QY 198 catgctgcttctgagaaagttttctgctcagagattggaccagaagcagtaacgctt 257  
 Db 1913 CTTGTGTCTTCTGAACCTTTGTCGGGCTTGATTGTTTGCATCAGCAGCTGAGCTTGTG 1854  
 QY 258 ttttggccatattcacacagataaaatgaacacataattgaaatgctccacacccctgtg 317  
 Db 1853 TTGTGGGATAGTTTCTACAGATTTTGTCTTAAATATATATAAGACCGATAGATGTTA 1794  
 QY 318 gccattgttgatcttccgggaatgatccatcccttcttatccatctatcgcacatcc 377  
 Db 1793 TCCATCCTAGTCGCTTCAATAGTCCTTTTGTGCTCTCATAGCTTGAGATTTCAACAT 1734  
 QY 378 tcaagatttggaaaactgttgggaagtggcagctgaacacttctatgga 427  
 Db 1733 TTAAGCTTTGAAATGCATTTTCGGATATGTTTAAATCTTTTCTCTGTA 1684

RESULT 8

US-08-471-733-7/c  
 ; Sequence 7, Application US/08471733  
 ; Patent No. 6068842  
 GENERAL INFORMATION:  
 APPLICANT: BERGSTROM SVEN  
 APPLICANT: BARBOUR ALAN G.  
 TITLE OF INVENTION: NEW 66 KDA ANTIGEN FROM BORRELIA  
 NUMBER OF SEQUENCES: 14  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: BROWDY AND NEIMARK  
 STREET: 419 SEVENTH STREET, N.W.  
 CITY: WASHINGTON  
 COUNTRY: USA  
 ZIP: 20004  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/471,733  
 FILING DATE: 06-JUN-1995  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/262,220  
 FILING DATE: 20-JUN-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: COOPER, IVER P.  
 REGISTRATION NUMBER: 28,005  
 REFERENCE/DOCKET NUMBER: BERGSTROM-3  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 202-628-5197  
 TELEFAX: 202-737-3528  
 TELEX: 248633  
 INFORMATION FOR SEQ ID NO: 7:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2547 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 ORIGINAL SOURCE:  
 ORGANISM: Borrelia garinii  
 STRAIN: Ip90  
 IMMEDIATE SOURCE:  
 CLONE: PJB-101  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 380..2245  
 FEATURE:

NAME/KEY: sig\_peptide  
LOCATION: 380..442  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 443..2242  
US-08-471-733-7

Query Match 2.58; Score 36.4; DB 3; Length 2547;  
Best Local Similarity 47.4%; Pred. No. 0.47;  
Matches 109; Conservative 0; Mismatches 121; Indels 0; Gaps 0;  
QY 198 catggtcttcctgagaagttttcttcagagattgacacagcagtaacgctt 257  
DB 1913 CTTGTGCTCCGAACTTTCGCGCTTGATTTGTCATCAGCAGCTGAGCTTTG 1854  
QY 258 ttttgggcattatcacacgataaataatgaaatgtcccaacacctcgtg 317  
DB 1853 TTGTGGGATAGTTTCTACAGATTTTGTCTTAAATATCATAAAGACCGATAGATGTTA 1794  
QY 318 gcaatgttgattccgggaatgatccatccctttcttaccactactatcgccatcc 377  
DB 1793 TCCATCTAGTCCGTTTCAATAGTCCCTTTTGTGTTGTTCTTATAGCTTGAGATTCAACAT 1734  
QY 378 tcaaggatttggaaactgttggaagtgcagctgaaacacttctatgga 427  
DB 1733 TTAAGCTTTGAATGCAATTTTCGGATAGTGTTTTAAATCTTTTCTGTGA 1684

RESULT 9  
US-08-468-878-7/c  
; Sequence 7, Application US/08468878  
; Patent No. 6090586  
; GENERAL INFORMATION:  
; APPLICANT: BERGSTROM SVEN  
; APPLICANT: BARBOUR ALAN G.  
; TITLE OF INVENTION: NEW 66 KDA ANTIGEN FROM BORRELIA  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 SEVENTH STREET, N.W.  
; CITY: WASHINGTON  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/468,878  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/262,220  
; FILING DATE: 20-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: COOPER, IVER P.  
; REGISTRATION NUMBER: 28,005  
; REFERENCE/DOCKET NUMBER: BERGSTROM-3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; TELEX: 248633  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2547 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; ORIGINAL SOURCE:  
; ORGANISM: Borrelia garinii

STRAIN: Ip90  
IMMEDIATE SOURCE:  
CLONE: PJB-101  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 380..2245  
FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: 380..442  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 443..2242  
US-08-468-878-7

Query Match 2.58; Score 36.4; DB 3; Length 2547;  
Best Local Similarity 47.4%; Pred. No. 0.47;  
Matches 109; Conservative 0; Mismatches 121; Indels 0; Gaps 0;  
QY 198 catggtcttcctgagaagttttcttcagagattgacacagcagtaacgctt 257  
DB 1913 CTTGTGCTCCGAACTTTCGCGCTTGATTTGTCATCAGCAGCTGAGCTTTG 1854  
QY 258 ttttgggcattatcacacgataaataatgaaatgtcccaacacctcgtg 317  
DB 1853 TTGTGGGATAGTTTCTACAGATTTTGTCTTAAATATCATAAAGACCGATAGATGTTA 1794  
QY 318 gcaatgttgattccgggaatgatccatccctttcttaccactactatcgccatcc 377  
DB 1793 TCCATCTAGTCCGTTTCAATAGTCCCTTTTGTGTTGTTCTTATAGCTTGAGATTCAACAT 1734  
QY 378 tcaaggatttggaaactgttggaagtgcagctgaaacacttctatgga 427  
DB 1733 TTAAGCTTTGAATGCAATTTTCGGATAGTGTTTTAAATCTTTTCTGTGA 1684

RESULT 10  
US-08-750-494-7/c  
; Sequence 7, Application US/08750494  
; Patent No. 6204018  
; GENERAL INFORMATION:  
; APPLICANT: BERGSTROM SVEN  
; APPLICANT: BARBOUR ALAN G.  
; TITLE OF INVENTION: NEW 66 KDA ANTIGEN FROM BORRELIA  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 SEVENTH STREET, N.W.  
; CITY: WASHINGTON  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/750,494  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/262,220  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: COOPER, IVER P.  
; REGISTRATION NUMBER: 28,005  
; REFERENCE/DOCKET NUMBER: BERGSTROM-3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; TELEX: 248633  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:

```

; LENGTH: 2547 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Borrelia garinii
; STRAIN: IP90
; IMMEDIATE SOURCE:
; CLONE: pJB-101
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 380..2245
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 380..442
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 443..2242
; US-08-750-494-7

Query Match          2.5%; Score 36.4; DB 4; Length 2547;
Best Local Similarity 47.4%; Pred. No. 0.47;
Matches 109; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 198 catggcgcttcctgaagaagttttctgcttcagagattgaccagaagcagtaacggctt 257
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1913 CTGTGCTGCTCTCTGAACCTTGTCCGGGCTGATTTGGTCATCAGCAGCTTGCTTG 1854
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 258 ttttggcgcatattccaaacgataaaatgaacataattgaaaatgctcccaacacctcgtg 317
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1853 TTGTGGGATAGTTTTCTACAGATTTTTTGTCTTAAATATATATAAGACCGATAGATGTTA 1794
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 318 gccatgttgatctctccgggaatgatccatcccttcttccactactcatcgccatcc 377
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1793 TCCATCCTCTAGTCGGTTCATAAGTCCTTTTTTTGGTGTCTTCATAGCTTGAGATTTCACAT 1734
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 378 tcaagatttgaaactgtgtggaagtggcagctgaacactcttatgga 427
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1733 TTAAGCTTTGAATGCATTTTCGGATATGGTTTTTAATCTTTTCTGTGTA 1684
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
US-08-471-717-1
; Sequence 1, Application US/08471717
; Patent No. 5859337
; GENERAL INFORMATION:
; APPLICANT: Gasser, Charles S.
; APPLICANT: Lippuner, Veronica
; TITLE OF INVENTION: GENES CONFERRING SALT TOLERANCE AND
; TITLE OF INVENTION: THEIR USES
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew
; STREET: One Market Plaza, Steuart Street Tower
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105-1492
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,717
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 2307E-606

```



RESULT 15  
 US-09-150-741-1/c  
 ; Sequence 1, Application US/09150741  
 ; Patent No. 6183996  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Stewart et al.  
 ; TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate  
 ; Patent No. 6183996  
 ; TITLE OF INVENTION: Synthetase II  
 ; FILE REFERENCE:  
 ; CURRENT APPLICATION NUMBER: US/09/150,741  
 ; CURRENT FILING DATE: 1998-09-10  
 ; EARLIER APPLICATION NUMBER: PL6380  
 ; EARLIER FILING DATE: 1992-12-16  
 ; EARLIER APPLICATION NUMBER: AU93/00617  
 ; EARLIER FILING DATE: 1993-12-02  
 ; EARLIER APPLICATION NUMBER: 08/446,855  
 ; EARLIER FILING DATE: 1995-07-06  
 ; NUMBER OF SEQ ID NOS: 15  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 1  
 ; LENGTH: 8920  
 ; TYPE: DNA  
 ; ORGANISM: Plasmodium falciparum  
 US-09-150-741-1

Query Match 2.4%; Score 35.6; DB 4; Length 8920;  
 Best Local Similarity 52.7%; Pred. No. 1.6;  
 Matches 77; Conservative 0; Mismatches 69; Indels 0; Gaps 0;  
 Qy 1338 ttgttgtagaaattgtgttattgcaacaggtagagaaacataaccatagacagatgtatc 1397  
 Db 401 TATATTAATAAGTTTATATTTTCAAGTAATTTTATAACAAATGAACACACAAACATATA 342  
 Qy 1398 tgaagagataagcttctctatgtctaaagaaatggaccgatacgaataaaaaaacatca 1457  
 Db 341 TATATATATATATATATATATATATATATATATATAATAAATAACTTAAATGTATTGTTA 282  
 Qy 1458 ttaagattaaaaaaataaaaaaa 1483  
 Db 281 ATAAATATAAGAAAAAATAAAAAA 256

Search completed: July 12, 2001, 01:48:19  
 Job time: 5379 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2001, 00:12:55 ; Search time 1368.44 Seconds  
(without alignments)  
10244.206 Million cell updates/sec

Title: US-09-545-072A-1  
Perfect score: 1483  
Sequence: 1 attgcaaccaggagagaaaa.....attaaaaaa 1483

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

1:	gb_est1.*
2:	gb_est2.*
3:	gb_est3.*
4:	gb_est4.*
5:	gb_est5.*
6:	gb_est6.*
7:	gb_est7.*
8:	gb_est8.*
9:	gb_est9.*
10:	gb_est10.*
11:	gb_est11.*
12:	gb_est12.*
13:	gb_est13.*
14:	gb_est14.*
15:	gb_est15.*
16:	gb_est16.*
17:	gb_est17.*
18:	gb_est18.*
19:	gb_est19.*
20:	gb_est20.*
21:	gb_est21.*
22:	gb_est22.*
23:	gb_est23.*
24:	gb_est24.*
25:	gb_est33.*
26:	gb_est34.*
27:	gb_est35.*
28:	gb_est36.*
29:	gb_est37.*
30:	gb_est38.*
31:	gb_est39.*
32:	gb_est40.*
33:	em_estba.*
34:	em_estfun.*
35:	em_esthum1.*
36:	em_esthum2.*
37:	em_esthum3.*
38:	em_esthum4.*
39:	em_esthum5.*
40:	em_esthum6.*
41:	em_esthum7.*
42:	em_esthum8.*
43:	em_esthum9.*

44:	em_esthum10.*
45:	em_esthum11.*
46:	em_esthum12.*
47:	em_esthum13.*
48:	em_esthum14.*
49:	em_esthum15.*
50:	em_esthum16.*
51:	em_esthum17.*
52:	em_esthum18.*
53:	em_esthum19.*
54:	em_esthum20.*
55:	em_esthum21.*
56:	em_esthum22.*
57:	em_esthum23.*
58:	em_esthum24.*
59:	em_esthum25.*
60:	em_esthum26.*
61:	em_esthum27.*
62:	em_esthum28.*
63:	em_estin1.*
64:	em_estin2.*
65:	em_estin3.*
66:	em_estin4.*
67:	em_estin5.*
68:	em_estom1.*
69:	em_estom2.*
70:	em_estov1.*
71:	em_estov2.*
72:	em_estpl1.*
73:	em_estpl2.*
74:	em_estpl3.*
75:	em_estpl4.*
76:	em_estpl5.*
77:	em_estpl6.*
78:	em_estpl7.*
79:	em_estpl8.*
80:	em_estpl9.*
81:	em_estpl10.*
82:	em_estrol.*
83:	em_estrol2.*
84:	em_estrol3.*
85:	em_estrol4.*
86:	em_estrol5.*
87:	em_estrol6.*
88:	em_estrol7.*
89:	em_estrol8.*
90:	em_estrol9.*
91:	em_estrol10.*
92:	em_estrol11.*
93:	em_estrol12.*
94:	em_estrol13.*
95:	em_estrol14.*
96:	em_estrol15.*
97:	em_estrol16.*
98:	em_estrol17.*
99:	em_estrol18.*
100:	em_estrol19.*
101:	em_estrol20.*
102:	gb_est25.*
103:	gb_est26.*
104:	gb_est27.*
105:	gb_est28.*
106:	gb_est29.*
107:	gb_est30.*
108:	gb_est31.*
109:	gb_est32.*
110:	gb_est41.*
111:	gb_est42.*
112:	gb_est43.*
113:	gb_est44.*
114:	gb_est45.*
115:	gb_est46.*
116:	gb_est47.*

117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: gb\_est121:\*  
191: gb\_est122:\*  
192: gb\_est123:\*  
193: gb\_est124:\*  
194: gb\_est125:\*  
195: gb\_est126:\*  
196: gb\_est127:\*  
197: gb\_est128:\*  
198: gb\_est129:\*  
199: gb\_est130:\*  
200: gb\_est131:\*  
201: gb\_est132:\*  
202: gb\_est133:\*  
203: gb\_est134:\*  
204: gb\_est135:\*  
205: gb\_est136:\*  
206: gb\_est137:\*  
207: gb\_est138:\*  
208: gb\_est139:\*  
209: gb\_est140:\*  
210: gb\_est141:\*  
211: gb\_est142:\*  
212: gb\_est143:\*  
213: gb\_est144:\*  
214: gb\_est145:\*  
215: gb\_est146:\*  
216: gb\_est147:\*  
217: gb\_est148:\*  
218: gb\_est149:\*  
219: gb\_est150:\*  
220: gb\_est151:\*  
221: gb\_est152:\*  
222: gb\_est153:\*  
223: gb\_est154:\*  
224: gb\_est155:\*  
225: gb\_est156:\*  
226: gb\_est157:\*  
227: gb\_est158:\*  
228: gb\_est159:\*  
229: gb\_est160:\*  
230: gb\_est161:\*  
231: gb\_est162:\*  
232: gb\_est163:\*  
233: gb\_est164:\*  
234: gb\_est165:\*  
235: gb\_est166:\*  
236: gb\_est167:\*  
237: gb\_est168:\*  
238: gb\_est169:\*  
239: gb\_est170:\*  
240: gb\_est171:\*  
241: gb\_est172:\*  
242: gb\_est173:\*  
243: gb\_est174:\*  
244: gb\_est175:\*  
245: gb\_est176:\*  
246: gb\_est177:\*  
247: gb\_est178:\*  
248: gb\_est179:\*  
249: gb\_est180:\*  
250: gb\_est181:\*  
251: gb\_est182:\*  
252: gb\_est183:\*  
253: gb\_est184:\*  
254: gb\_est185:\*  
255: gb\_est186:\*  
256: gb\_est187:\*  
257: gb\_est188:\*  
258: gb\_est189:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.



## SUMMARIES

Result No.	Score	Query		Length	DB ID	Description
		Match	Count			
C	1	334.2	22.5	347	30	AV521636
C	2	236.2	15.9	253	136	BE528478
C	3	185	12.5	541	21	AI486368
C	4	176.6	11.9	455	115	AW395171
C	5	128.4	8.7	598	146	BF258527
C	6	117.2	7.9	821	146	BF267473
C	7	115.4	7.8	505	119	AW670546
C	8	98.4	6.6	626	23	AI670146
C	9	93.4	6.3	440	136	BE500171
C	10	91	5.9	605	112	AW172078
C	11	88	5.9	432	155	BG558083
C	12	82.6	5.6	480	143	BF053864
C	13	79.2	5.3	532	173	BG098367
C	14	78.2	5.3	482	154	BG465365
C	15	72.4	4.9	486	173	BG098731
C	16	66.4	4.5	523	256	BG27915
C	17	63.2	4.3	896	152	BG369497
C	18	63	4.2	381	174	BG135784
C	19	55.8	3.8	191	256	B76951
C	20	49	3.3	1101	219	CNS00319R
C	21	48.6	3.3	928	219	CNS00DKY
C	22	46.6	3.1	324	112	AW164040
C	23	46.6	3.1	353	165	BES232178
C	24	46.6	3.1	1099	219	CNS00DF9
C	25	45.8	3.1	945	221	CNS04D0K
C	26	45.8	3.1	965	219	CNS0056E
C	27	45.6	3.1	408	105	AI514793
C	28	45.4	3.1	419	21	A1536685
C	29	45.4	3.1	890	219	CNS00AFU
C	30	45	3.0	220	107	AU057778
C	31	44.8	3.0	238	221	CNS04CWO
C	32	44.8	3.0	311	105	AI513763
C	33	44.8	3.0	480	141	BE881315
C	34	44.8	3.0	1101	219	CNS012TP
C	35	44.6	3.0	503	105	AI513809
C	36	44.6	3.0	938	219	CNS006TU
C	37	44.2	3.0	374	139	BE754762
C	38	44.2	3.0	376	139	BE752168
C	39	44.2	3.0	691	32	AV729259
C	40	44	3.0	374	23	A1671679
C	41	44	3.0	646	32	AV682770
C	42	44	3.0	831	220	CNS023K1
C	43	44	3.0	862	245	AZ539974
C	44	44	3.0	1201	219	CNS016EA
C	45	43.8	3.0	299	107	AU035002

## ALIGNMENTS

RESULT	1	
AV521636/c		
LOCUS	AV521636	
DEFINITION	347 bp mRNA EST	07-SEP-2000
	Arabidopsis thaliana aboveground organs two to six-week old	
	Arabidopsis thaliana CDNA clone AP263b12F 3', mRNA sequence.	
ACCESSION	AV521636	
VERSION	AV521636.1	
KEYWORDS	GI:8691163	
SOURCE	EST.	
ORGANISM	thale cress.	
	Arabidopsis thaliana	
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.	
REFERENCE	(bases 1 to 347)	
AUTHORS	Asanizu,E., Nakamura,Y., Sato,S. and Tabata,S.	
TITLE	A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized and	







Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoideae; Andropogoneae; Zea.  
1. (bases 1 to 626)  
Walbot, V.  
Maize ESTs from various cDNA libraries sequenced at Stanford  
University  
Unpublished (1999)  
Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Plate: 605019 row: A column: 04.

FEATURES  
source0  
0  
4  
3  
0  
2

BASE COUNT	Query Match	
ORIGIN	Best Local	
	Matches	
	Qy	1020
	Db	536

Db	476	CT
Ov	1140	ET

Db	416	TG
Qy	1200	999
Db	356	GA
RESULT	9	
BE500171		
LOCUS		
DEFINITION		
ACCESSION		
VERSION		
KEYWORDS		

ORGANISM	
REFERENCE AUTHORS	
TITLE	
JOURNAL COMMENT	

JOURNAL  
COMMENT

US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center  
800 Buchanan Street, Albany, CA 94710, USA  
Tel: 5105595773  
Fax: 5105595818  
Email: oandersn@pw.usda.gov

Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20  
Seq primer: Stratagene SK primer.

#### FEATURES

source  
1..440  
/organism="Triticum aestivum"  
/cultivar="Chinese Spring"  
/db\_xref="taxon:4565"  
/clone="WHE0980.G12.M24"  
/clone\_lib="wheat pre-anthesis spike cDNA library"  
/tissue\_type="Spike before anthesis"  
/dev\_stage="Adult plant"  
/lab\_host="E. coli SOLR"  
/note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site 1: EcoRI; Site 2: XhoI; plants were grown in the greenhouse. Whole spike with awns trimmed, white, green and yellow anther were collected and total RNA, and poly(A) RNA were prepared, a cDNA library was made, and the cDNA clones were in vivo excised to give pBluescript phagemids in the TJ Close lab (Choi, Close, Fenton) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

BASE COUNT 81 a 152 c 113 g 94 t

Query Match 6.3%; Score 93.4; DB 136; Length 440;

Best Local Similarity 59.3%; Pred. No. 1e-14;

Matches 182; Conservative 0; Mismatches 116; Indels 9; Gaps 1;

QY 112 ctgaagctcaatgaagcttataagcattgggtttggagaaatagagatgtatgaacatc 171

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 143 CCGAGTCGCCATGAGGCTACAGAGTCTGGTGGGAGAACCGGACCTCGCTCGGCTC 202

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 172 cttggatcctttcccaacgattgacatgctgctctccctcgagaagttttctcctcaga 231

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 203 CCTGAGTCCCTCGCCCAACGGGGTGACATGGATATCTCTGAGCGCTTCGCTAACCCGA 262

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 232 gattggaccagaagcagtaacgcttttttgggcattattcacaacgataaataaacacat 291

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 263 GATTGCCCGGACGATATATGACATCTCTGGGCATTTGAATGTCCTCAACGACATAT 322

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 292 aattgaaaatgctcccaacacctcgtggccatgttggatcttcocgggaatgatccatccct 351

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 323 AATTGA-----GACACCAACTGATGGTCTACTACTGCGCTCCCAAGGAACAATCTAT 373

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 352 ttcttatccactactccatccatccagatttgagaaactgttgtgaagtgcagc 411

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 374 CCCATGGGCTCTTGTGATCTATACTCAAGGATGTCGAAGCAGTGTGTAAGTTGCCGC 433

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 412 tgaacac 418

|||||

Db 434 CCAGCAC 440

|||||

RESULT 10

AW172078/c

LOCUS AW172078 605 bp mRNA EST 15-NOV-1999

DEFINITION 618049D09.y1 618 - Inbred Tassel cDNA Library Zea mays cDNA, mRNA

sequence.

ACCESSION AW172078

VERSION AW172078.1 GI:6431874

KEYWORDS EST.

SOURCE Zea mays.

ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 605)

Maize ESTs from various cDNA libraries sequenced at Stanford

University

Unpublished (1999)

Contact: Walbot V

Department of Biological Sciences

Stanford University

855 California Ave, Palo Alto, CA 94304, USA

Tel: 650 723 2227

Fax: 650 725 8221

Email: walbot@stanford.edu

Plate: 618049 row: D column: 09.

Location/Qualifiers

1..605

/organism="Zea mays"

/cultivar="Ohio43"

/db\_xref="taxon:4577"

/clone\_lib="618 - Inbred Tassel cDNA Library"

/tissue\_type="tassel"

/dev\_stage="tassel length from 0.1 to 2.5 cm"

/lab\_host="XLOLR"

/note="Organ: tassel; Vector: pAD-GAL4-2.1 (Hybrizap);

Inbred tassel library from Schmidt lab"

BASE COUNT 191 a 150 c 118 g 146 t

ORIGIN

Query Match 6.1%; Score 91; DB 112; Length 605;

Best Local Similarity 66.7%; Pred. No. 5e-14;

Matches 130; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 1029 atgagctgagagacgaaactgatatggcatttacctcatgagatccattcttca 1088

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 472 ATCAGTTGAAAAGGCCAAAGATGATGTGGGCTCTTATGTGAGAGATCTTCTTTG 413

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1089 ccaagtacacagagcagaagctggaagctctcaaaagaagctggaacttaattccattga 1148

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 412 CCAGTTACAGCAAGGCTCACCTCTCGAAGGCTGAACAGGTTCTGAATCCGGTGCCATTGA 353

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1149 tcgagattctccacagagaagattgtggaagcttttggaggagctcagtcacggtcacatt 1208

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 352 TTGGCTTCCTTACAGGAACTTGTAGAGCTACTGGAGGGGATTGACAGGATACACGT 293

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1209 acatatcgggatcgt 1223

||||| ||||| |||||

Db 292 ACACATCAGGTTTCAT 278

||||| ||||| |||||

RESULT 11

BG558083

LOCUS RH122\_66\_E03.g1\_A003 Rhizome2 (RH122) Sorghum propinquum cDNA, mRNA

DEFINITION sequence.

ACCESSION BG558083

VERSION BG558083.1 GI:13587081

KEYWORDS EST.

SOURCE Sorghum propinquum.

ORGANISM Sorghum propinquum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC

clade; Panicoideae; Andropogoneae; Sorghum.

1 (bases 1 to 432)

Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt

,L.H.

An EST database from Sorghum: Sorghum propinquum rhizomes

Unpublished (2000)

Contact: Cordonnier-Pratt MM

Department of Botany

The University of Georgia

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

```
/lab_host="SOLR"
/notes=vector: pBlueScript SK(-); Site_1: EcoRI; Site_2:
XhoI; Tissue was supplied by Dr. Fry (Cornell University).
Leaflets and petioles were isolated from 8 week old
greenhouse grown plants. The plants were watered and
fertilized freely. The tissue was immediately frozen in
liquid nitrogen."
```

BASE COUNT	136 a	79 c	107 g	158 t
ORIGIN				

	Query Match	5.6%	Score 82.6	DB 143	Length 480
Best Local Similarity	58.4%	Prod No	9	4e-12	
Matches 184	Conservative	0	Mismatches 124	Indels 7	Gaps 2
Qy 122	atggaagcttataagcaatgggtttgagaataagagagatgtacaaatccctttggatccc	181			

163 ATGGAGGGTTTATAAGAGAGATGGATTAGGAGGGAATAAGAG

[illegible]

283 AGAAGCAGTGCATCCCTTGAGGAATGATTCTACCGTGACGAATATATAATAGAAC 342

343 AAGTCCAAC-----CGAGGTGCATACGAGATGTGCAGAGTCTCTTTATCCCTTTATC 396  
 Db

D<sub>b</sub> 397 ATTGTGCCTTACCTTGCTGTAAGACTTTGGAAACATCAATTGAGGTGTGGCTGAGCAAAT 456

421	ctatcgagcacaacaa	455	ctatcgagcacaacaa
457	TTATGGTGAAGAA	471	TTATGGTGAAGAA

RESULT	13
BG098367	
LOCUS	532 bp mRNA
DEFINITION	EST452886 sprouting eyes/shoots Solanum tuberosum cDNA clone
	CSTC2U23 5' sequence, mRNA sequence.
ACCESSION	BG098367

VERSION	EG036367.1	GI:12368402
KEYWORDS		
SOURCE	potato.	
ORGANISM	Solanum tuberosum	

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.

1 (bases 1 to 532)

REFERENCE

**AUTHORS** van der Hoeven, R.S., Bezzerides, J., Cho, J., Utterback, T., Hansen, C.L., Bougri, O., Buell, C.R., Ronning, C.M., Tanksley, S.D., and Baker, B.

**TITLE** Generation of ESTs from potato sprouting eyes/shoots

**JOURNAL** Unpublished (2001)

**COMMENT** Contact: Cathy Ronning

**CONTACT:** Kathy Manning  
The Institute for Genomic Research  
For clone info: please contact Research Genetics, Libraries  
Division tel 1-800-711-6195, email [cdna@resgen.com](mailto:cdna@resgen.com).  
Location/Qualifiers

```

1. 532
/organism="Solanum tuberosum"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="CSR2J23"
/clone_lib="sprouting eyes/shoots"
/tissue_type="sprouting tubers"
/dev_stage="12-14 weeks post harvest"
/lab_host="SOLR"

```





	Matches	94;	Conservative	0;	Mismatches	36;	Indels	0;	Gaps	0;
Qy	119	tcaatggaagccttataagcaaatgggttttgagaaatagagagatgtatgacaatcctttgga	178							
Db	211	TCTATGGAGGCTTACAAGAGATGGTCCGAAGGAATCGTGATTATGTCGCCCAATTGAGT	270							
Qy	179	tcctttgccaaagattgacatggctgcttcctgagaaagtttctgcttcagagattgga	238							
Db	271	TCTCTAGCCAGTGGGAATGACATGGCTTCTCCAGAGCGGTTTGCTGCATCAGAAATCGGG	330							
Qy	239	ccagaagcag	248							
Db	331	CCAGAAGCAG	340							

Search completed: July 12, 2001, 01:12:04  
Job time: 3549 sec



GenCore version 4.5  
 Copyright (c) 1993 - 2000 CompuGen Ltd.  
 OW protein - protein search, using sw model  
 Run on: July 11, 2001, 11:35:01 ; Search time 13.09 Seconds  
 (without alignments)  
 960.410 Million cell updates/sec  
 Title: US-09-545-072A-2  
 Perfect score: 1922  
 Sequence: 1 MEAYKQWVRNREYVQSFSGS.....EKIVELLEGAQSRVITYISGS 367  
 Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5  
 Searched: 93435 seqs, 34255486 residues  
 Total number of hits satisfying chosen parameters: 93435  
 Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000  
 Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries  
 Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	186.5	9.7	391	1	PEXG_YARLI
2	96.5	5.0	4486	1	DYH9_HUMAN
3	96	5.0	987	1	YD94_METJA
4	93	4.8	1306	1	MSB2_YEAST
5	89	4.6	549	1	DSX_DROME
6	88	4.6	1706	1	CYAA_BORPE
7	86.5	4.5	4466	1	DYHC_ANTCR
8	85.5	4.4	758	1	CHEA_RHIME
9	85.5	4.4	879	1	MCW3_SCHPO
10	85	4.4	248	1	AMPB_BACHD
11	84	4.4	3411	1	POLG_YEFV1
12	84	4.4	3411	1	POLG_YEFV2
13	83.5	4.3	1325	1	Y309_MYCPN
14	83	4.3	524	1	CP72_CATRO
15	83	4.3	1142	1	ENAM_PIG
16	82.5	4.3	2145	1	CYAA_PODAN
17	82	4.3	566	1	YJ19_SYNY3
18	82	4.3	4829	1	BIR6_HUMAN
19	81.5	4.2	897	1	APG1_YEAST
20	81.5	4.2	1250	1	SSD1_YEAST
21	81	4.2	957	1	KP2_KLULA
22	80.5	4.2	248	1	AMPB_BACSU
23	80.5	4.2	467	1	IRF6_HUMAN
24	80.5	4.2	524	1	CK13_YEAST
25	80.5	4.2	752	1	ABC7_HUMAN
26	79.5	4.1	474	1	CISY_EMEVI
27	79.5	4.1	531	1	UL95_HCMVA
28	79.5	4.1	637	1	YHES_ECOLI
29	79.5	4.1	821	1	CTRL_ARATH
30	79.5	4.1	896	1	APCE_SYNY4
31	79.5	4.1	911	1	YJ20_MOUSE
32	79.5	4.1	1189	1	YJH6_YEAST
33	79.5	4.1	2076	1	FAS1_YARLI

34 79.5 4.1 2442 1 CBP\_HUMAN  
 35 79.5 4.1 4466 1 DYHC\_TRIGR  
 36 79 4.1 380 1 BCAL\_SCHPO  
 37 79 4.1 590 1 OAM\_ASCSU  
 38 79 4.1 598 1 Y427\_HUMAN  
 39 78.5 4.1 1288 1 VACA\_HELPJ  
 40 78.5 4.1 2194 1 SC16\_YEAST  
 41 78.5 4.1 4273 1 PKSM\_BACSU  
 42 78 4.1 176 1 VLPA\_MYCHR  
 43 78 4.1 427 1 YF09\_MYCPN  
 44 78 4.1 435 1 GAT3\_XENLA  
 45 78 4.1 518 1 KR2\_HSV11

## ALIGNMENTS

RESULT 1  
 PEXG\_YARLI  
 ID PEXG\_YARLI STANDARD; PRT; 391 AA.  
 AC P78980;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE PEROXISOMAL MEMBRANE PROTEIN PEX16 (PEROXIN-16).  
 GN PEX16.  
 OS Yarrowia lipolytica (Candida lipolytica).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Dipodascaceae; Yarrowia.  
 OX NCBI\_TaxID=4952;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=E122;  
 RX MEDLINE=97327755; PubMed=9182661;  
 RA Eitzen G.A., Szillard R.K., Rachubinski R.A.;  
 RT "Enlarged peroxisomes are present in oleic acid-grown Yarrowia  
 lipolytica overexpressing the PEX16 gene encoding an intraperoxisomal  
 peripheral membrane peroxin";  
 RL J. Cell Biol. 137:1265-1278(1997).  
 CC -!- SUBCELLULAR LOCATION: PEROXISOMAL MEMBRANE-ASSOCIATED.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; U75433; ABA41724.1; -.  
 KW Peroxisome; Membrane.  
 SQ SEQUENCE 391 AA; 44479 MW; F328E3C6CEFC655 CRC64;

Query Match 9.7%; Score 186.5; DB 1; Length 391;  
 Best Local Similarity 20.7%; Pred. No. 8.5e-09;  
 Matches 86; Conservative 68; Mismatches 143; Indels 119; Gaps 13;  
 Oy 1 MEAYKQWVRNREYVQSFSGFANGTLWLLPKFSASEIGPEAVTAFGLGFTTINEHIEN 60  
 Db 20 LDSYDKFLVRNNAASGISTESTLRVSVYLPORFNDVEIATETLYAVNLVGLYHDTIAR 79  
 Oy 61 APTPRGHVSGNDSPLSYSP-----LLIAILKDLFTVVVEVAAEH 99  
 Db 80 A-----VAASPNAAAVYRPSPHNRYTDWFTKRNKGYKASRAVTFVKFGLVAEMVAKK 133  
 Oy 100 FYGD-KKNYIILTEAMKAVIRLALFRNSGYKMLLOGGETPNEEK----- 143  
 Db 134 NGGEMARWKCITIGETIKAGRIYMLGSTLYQPLC---TTPYDPREVTGELLETICRDEG 190  
 Oy 144 ---DSNQSESNRAGNSGNLPGHGLGNQHNHPNLEGRAMSALSSFGONARTTTTSS 199  
 Db 191 ELDIKGLMDPQWKMPRTGRTIPEIA-----PTNVEGYLLT----- 226

```
QY 200 PWSRRIOHQAVIEPPMIKERRTMSLTKGVNCAIPALGEVLYITRPLIYVLFIRK 259
Db 227 -----KVLRSVDPRPNLLSR-----LDNMGV-----VAELSLIRPLIYACLIFR 268
QY 260 YGVR-----WIPWALSISVDTLGMLLANSKWKGEKSKQVHFGSP--- 300
Db 269 QHVNKTPASTKSKFPLNSPAPWIIIGLVIEALSRKMWS-----WLLRQO---SCKTPT 322
QY 301 --EKDELRRKLIALYLMRDPFTTKYTRQKLESSOKKLELPIGLFLEKIVELL 354
Db 323 ALQOMEVKGRNLLGWLFRGEFYQATYRPLYSIVARLEKIPGLGFGALISDYL 378

RESULT 2
DIH9_HUMAN STANDARD; PRT; 4486 AA.
AC Q9N028; Q9N028; O95494;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CYLIARY DYNEIN HEAVY CHAIN (AXONEMAL DYNEIN HEAVY CHAIN) (DYNEIN
DE HEAVY CHAIN 9).
GN DHA9 OR DHAH17L OR DNALI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Reed W., Moats-Staats B.M., Carson J.L., Leigh M.W., Collier A.M.;
RA "A ciliary dynein heavy chain whose expression is upregulated in
RA differentiating airway epithelium.";
RA Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Bartoloni L., Blouin J.L., Maiti A., Sainsbury A., Rossier C.,
RA Gehrig C., She J.X., Marron M.P., Lander E., Meeks M., Chung E.,
RA Jorissen M., Scott H.S., Delozier Blanchet C.D., Gardiner M.,
RA Antonarakis S.E.;
RA "Axonemal beta heavy chain dynein DHA9: cDNA sequence, genomic
RA structure and investigation of its role in primary ciliary
RA dyskinesia.";
RA Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1874-1974 FROM N.A.
RC TISSUE-Nasal polyps;
RA Maiti A.K., Mattel W.G., Jorissen M., Volz A., Ziegler A.,
RA Bouvagnet P.;
RA "Chromosomal localization of human dynein heavy chain genes.";
RA Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP CHARACTERIZATION.
RA MEDLINE=20558134; PubMed=11104725;
RA Reed W., Carson J.L., Moats-Staats B.M., Lucier T., Hu P.C.,
RA Brighton L., Gambing T.M., Huang C.H., Leigh M.W., Collier A.M.;
RA "Characterization of an axonemal dynein heavy chain expressed early in
RA airway epithelial cillogenesis.";
RA Am. J. Respir. Cell Mol. Biol. 23:734-741(2000).
CC -1- FUNCTION: FORCE GENERATING PROTEIN OF RESPIRATORY CILIA. PRODUCES
CC FORCE TOWARDS THE MINUS ENDS OF MICROTUBULES. DYNEIN HAS ATPASE
CC ACTIVITY.
CC -1- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF
CC INTERMEDIATE AND LIGHT CHAINS.
CC -1- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)
```

```
CC or send an email to license@isb-sib.ch.
CC EMBL; AF257737; AAF69004.1; -
DR EMBL; AJ404468; CAB94756.1; -
DR EMBL; AJ132088; CAA10561.1; -
DR MIM; 603338; -
KW Motor protein; Microtubules; Dynein; ATP-binding; Coiled coil.
FT DOMAIN 381 410 COILED COIL (POTENTIAL).
FT DOMAIN 504 529 COILED COIL (POTENTIAL).
FT DOMAIN 539 662 COILED COIL (POTENTIAL).
FT DOMAIN 752 823 COILED COIL (POTENTIAL).
FT DOMAIN 1326 1355 COILED COIL (POTENTIAL).
FT DOMAIN 3051 3154 COILED COIL (POTENTIAL).
FT DOMAIN 3285 3341 COILED COIL (POTENTIAL).
FT DOMAIN 3640 3675 COILED COIL (POTENTIAL).
FT NP_BIND 1870 1877 ATP (POTENTIAL).
FT NP_BIND 2151 2158 ATP (POTENTIAL).
FT NP_BIND 2478 2485 ATP (POTENTIAL).
FT NP_BIND 2825 2832 ATP (POTENTIAL).
FT CONFLICT 2505 2505 L -> V (IN REF. 2).
FT CONFLICT 3678 3678 T -> A (IN REF. 2).
FT CONFLICT 4374 4374 I -> M (IN REF. 2).
SQ SEQUENCE 4486 AA; 511927 MW; 996EDFDEB0B3EB1 CRC64;

Query Match 5.0%; Score 96.5; DB 1; Length 4486;
Best Local Similarity 21.1%; Pred. No. 17;
Matches 74; Conservative 44; Mismatches 136; Indels 97; Gaps 17;

QY 56 HIEENAPTRGH-----VGSGNDPSLSYPLLIALLKOLETVEVAAEHFY----- 101
Db 2809 HINRILESPRGNALLVGVGSGKQ---SLTRLAAFISSMD-VFQITLRKGYQIDFRKMDL 2864
QY 102 -----GDKWNYI-ILTEAMKAVIRLALFRNSCYKMLQGGTPEEKEKSNQSESONR 153
Db 2865 ASLCIKAGVKNLNTVFLMTDAQVADERFLYLND---LLASGEIPDLYSD---DEVENI 2917
QY 154 AGNSGRNLGPHGLGNHNPWNL-----EGRAMSALSFGQNARTTSTGP----- 201
Db 2918 ISNVRNEVKSQLVD-NRENCWKFFIDTRQLKVTLCFSPVGNKLRVRSRKFPAIVNCT 2976
QY 202 -----W-----SRIOHQAVIEPPMIKERRTMSLTKGVNCAIPALGEVLY 246
Db 2977 AIHWFEHPQQALESVLSRFLQNTGIEPTKQSIKSFMAFVHTSVNQTOSYLSNEORY 3036
QY 247 -ITRPLIYVLFIRKYGVRSWI---PWAISLSVDTLGMLLANSKWKGEKSKQVHFGSP 302
Db 3037 NYTPKSFLEFIRLY-QSLLHRRKELKCKTERLENGLL-----KLHSTSAQV 3083
QY 303 DELRRKKLWALYLMRDPFTTKYTRQKLESSOKKLE---LPIGLFLEKI 350
Db 3084 DDLK-----AKLAAQAEVELKQKNEADKLIQVGVGVTDKV 3118

RESULT 3
YD94_METJA STANDARD; PRT; 987 AA.
AC Q58789;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL PROTEIN MJ1394.
GN MJ1394.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
```





OS Bordetella pertussis.  
 CC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;  
 CC Bordetella.  
 OX NCBI\_taxid=520;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RA MEDLINE=88216178; PubMed=2897067;  
 RX STRAIN=18323;  
 RA Glaser P., Ladant D., Serer O., Pichot F., Ullmann A., Danchin A.;  
 RT "the calmodulin-sensitive adenylate cyclase of Bordetella pertussis;  
 RL cloning and expression in Escherichia coli.";  
 RL Mol. Microbiol. 2:19-30(1988).  
 RN [2]  
 RP SEQUENCE OF 1489-1706 FROM N.A., AND BIFUNCTIONAL PROTEIN DESCRIPTION.  
 RC STRAIN=18323;  
 RA MEDLINE=89091151; PubMed=2905265;  
 RX Glaser P., Sakamoto H., Bellalou J., Ullmann A., Danchin A.;  
 RT "Secretion of cyclolysin, the calmodulin-sensitive adenylate cyclase-  
 RL haemolysin bifunctional protein of Bordetella pertussis.";  
 RL EMBO J. 7:3997-4004(1988).  
 RN [3]  
 RP DOMAINS.  
 RA MEDLINE=91117021; PubMed=2007407;  
 RX Munier H., Gilles A.-M., Glaser P., Danchin A., Sarfati R., Barzu O.;  
 RA "Isolation and characterization of catalytic and calmodulin-binding  
 RT domains of Bordetella pertussis adenylate cyclase.";  
 RL Eur. J. Biochem. 196:469-474(1991).  
 RN [4]  
 RP MUTAGENESIS.  
 RA MEDLINE=89251630; PubMed=2542030;  
 RX Glaser P., Elmaoglou-Lazaridou A., Krin E., Ladant D., Barzu O.,  
 RA Danchin A.;  
 RT "Identification of residues essential for catalysis and binding of  
 RL calmodulin in Bordetella pertussis adenylate cyclase by site-directed  
 RL mutagenesis.";  
 RL EMBO J. 8:967-972(1989).  
 RN [5]  
 RP MUTAGENESIS.  
 RA MEDLINE=91266896; PubMed=2050107;  
 RX Glaser P., Munier H., Gilles A.-M., Krin E., Porumb T., Barzu O.,  
 RA Sarfati R., Pellicuer C., Danchin A.;  
 RT "Functional consequences of single amino acid substitutions in  
 RL calmodulin-activated adenylate cyclase of Bordetella pertussis.";  
 RL EMBO J. 10:1683-1688(1991).  
 RN [6]  
 RP REVIEW.  
 RA MEDLINE=93119764; PubMed=8418825;  
 RX Danchin A.;  
 RT "Phylogeny of adenylate cyclases.";  
 RL Adv. Second Messenger Phosphoprotein Res. 27:109-162(1993).  
 RN [7]  
 RP PALMITOYLATION AT LYS-983.  
 RA MEDLINE=95025937; PubMed=7939682;  
 RX Hackett M., Guo L., Shabanowitz J., Hunt D.F., Hewlett E.L.;  
 RT "Internal lysine palmitoylation in adenylate cyclase toxin from  
 RL Bordetella pertussis.";  
 RL Science 266:433-435(1994).  
 RN [8]  
 RP PALMITOYLATION AT LYS-860.  
 RA MEDLINE=99214144; PubMed=10196151;  
 RX Basar T., Havlicek V., Bezouskova S., Halada P., Hackett M., Sebo P.;  
 RT "The conserved lysine 860 in the additional fatty acylation site of  
 RL Bordetella pertussis adenylate cyclase is crucial for toxin function  
 RL independently of its acylation status.";  
 RL J. Biol. Chem. 274:10777-10783(1999).  
 CC -!- FUNCTION: THIS ADENYLATE CYCLASE BELONGS TO A SPECIAL CLASS OF  
 CC BACTERIAL TOXIN. IT CAUSES WHOOPING COUGH BY ACTING ON MAMMALIAN  
 CC CELLS BY ELEVATING CAMP-CONCENTRATION AND THUS DISRUPTS NORMAL  
 CC CELL FUNCTION.  
 CC -!- CATALYTIC ACTIVITY: ATP = 3',5'-CYCLIC AMP + PYROPHOSPHATE.  
 CC -!- ENZYME REGULATION: ACTIVATED BY HOST CALMODULIN.  
 CC -!- SUBCELLULAR LOCATION: SECRETED.  
 CC -!- DOMAIN: THE GLY-RICH REGION IS PROBABLY INVOLVED IN BINDING

CC CALCIUM, WHICH IS REQUIRED FOR TARGET CELL-BINDING OR CYTOLYTIC  
 CC ACTIVITY (BY SIMILARITY).  
 CC -!- PTM: RELEASED IN A PROCESSED FORM.  
 CC -!- PTM: PALMITOYLATED BY CYAC. THE TOXIN ONLY BECOMES ACTIVE WHEN  
 CC MODIFIED IN POSITION LYS-983.  
 CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO ADENYLATE CYCLASE  
 CC CLASS-2 FAMILY.  
 CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE RTX  
 CC PROKARYOTIC TOXIN FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; Y00545; CAA68613.1; -;  
 DR EMBL; X14199; CAA32411.1; -;  
 DR EMBL; A07292; CAA00653.1; -;  
 DR EMBL; A14850; CAA01202.1; -;  
 DR PIR; S00893; OYBRC.  
 DR PIR; S09403; S09403.  
 DR InterPro; IPR001343; -;  
 DR Pfam; PF00353; hemolysinCabin; 5.  
 DR PRINTS; PR00313; CABDNNGRPT.  
 DR PROSITE; PS00330; HEMOLYSIN\_CALCIIUM; 5.  
 KW Lyase; CAMP synthesis; ATP-binding; Hemolysis; Toxin; Virulence;  
 KW Whooping cough; Calcium-binding; Repeat; Lipoprotein; Palmitate.  
 FT CHAIN 1 312  
 FT CALMODULIN-SENSITIVE ADENYLATE  
 FT CYCLASE.  
 FT HEMOLYSIN (BY SIMILARITY TO E.COLI  
 FT HEMOLYSIN HLVA).  
 FT A, CATALYTIC.  
 FT B, ALA/GLY-RICH.  
 FT C, ASP/GLY-RICH.  
 FT D, ATP (POTENTIAL).  
 FT 28 X REPEATS, GLY-RICH.  
 FT 1.  
 FT 2.  
 FT 3.  
 FT 4.  
 FT 5.  
 FT 6.  
 FT 7.  
 FT 8.  
 FT 9.  
 FT 10.  
 FT 11.  
 FT 12.  
 FT 13.  
 FT 14.  
 FT 15.  
 FT 16.  
 FT 17.  
 FT 18.  
 FT 19.  
 FT 20.  
 FT 21.  
 FT 22.  
 FT 23.  
 FT 24.  
 FT 25.  
 FT 26.  
 FT 27.  
 FT 28.  
 FT PALMITATE.  
 FT PALMITATE.  
 FT D->E,N,Y,H: LOSS OF ACTIVITY.  
 FT D->N,Y,H: LOSS OF ACTIVITY.

FT MUTAGEN 298 H->R,P,L: LOSS OF ACTIVITY.  
 FT MUTAGEN 301 E->Q,K: LOSS OF ACTIVITY.  
 SQ SEQUENCE 1706 AA; 177506 MW; F00744524BDD442E CRC64;

Query Match 4.6%; Score 88; DB 1; Length 1706;  
 Best Local Similarity 21.7%; Pred. No. 25;  
 Matches 70; Conservative 39; Mismatches 104; Indels 110; Gaps 17;

QY 25 LTWLLPEKFSASIGPEAVTAF-----LGTFITNEHIENAPTPRGHV----- 68  
 :  
 Db 240 LLWKIARAGARSACVGFTEARRQRFRYDGMNIGVITDF-ELEVRNALNRRAHAVAQAQDVVQH 298  
 :  
 QY 69 GSSGNDPSLSYPLLIALLKDLETVEVAAEHYFGDKKNYIIILTEAMKAVIRLALFRNSG 128  
 :  
 Db 299 GTEQNNP---FP-----EADKK-----IFVWSA-----TG 320  
 :

QY 129 YKMLLOGG---ETPNEEKDSNOSEONRA-GNSGRNLGPHGLGNQHNHPWNLEGGRAMSA 184  
 :  
 Db 321 ESQMLTRGQLKEVGQGEYVFVENRAYGVAGSKSLFDDGLGA--PGVPSCR---- 372  
 :  
 QY 185 LSGFGQNARTSTSPGWRRIHQOAAVTEPMIKERRRTMSSELLTEKGNGALFAIGEY 244  
 :  
 Db 373 -SKFSDVLETPASPGLRR-----BSLGAVERQDSDGYSLDGVSRSFSLGEV 420  
 :  
 QY 245 -----LYTRPLIVLVLFIRKYGVRSWIPWAISLSVDTLGMLANSKWMEKSKQ 294  
 :  
 Db 421 SDMAAVEAELENTRQVLHA-----GARQ-----DDAEPGVGSASHWGQRALQ 464  
 :  
 QY 295 VHFSGPEKDELRRKLIALYLIM 317  
 :  
 Db 465 ---GAQAVAAAQR-LVHAIALM 482  
 :

RESULT 7  
 DYHC\_ANTCR STANDARD; PRT; 4466 AA.

ID AC P39057;  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE DYNEIN BETA CHAIN, CILIARY.  
 OS Anthocidaris crassipina (Sea urchin).  
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
 OC Echinoidea; Euechinoidea; Echinacea; Echinoidea; Echinometridae;  
 OC Anthocidaris.  
 OX NCBI\_TaxID=7629;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91326104; PubMed=1830928;  
 RA Ogawa K.;  
 RT "Four ATP-binding sites in the midregion of the beta heavy chain of dynein".  
 RL Nature 352:643-645(1991).  
 CC -1- FUNCTION: FORCE GENERATING PROTEIN OF EUKARYOTIC CILIA AND FLAGELLA. PRODUCES FORCE TOWARDS THE MINUS ENDS OF MICROTUBULES.  
 CC DYNELIN HAS ATPASE ACTIVITY.  
 CC -1- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS (ALPHA AND BETA), THREE INTERMEDIATE CHAINS AND SEVERAL LIGHT CHAINS.  
 CC -1- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to licensese@isb-sib.ch).

CC EMBL; DOI021; BAA00827.1; --  
 DR PIR; S17231; S17231.  
 DR Motor protein; Microtubules; Dynein; ATP-binding; Flagella;  
 KW Colled coil.



RT \*Analysis of a chemotaxis operon in Rhizobium meliloti.\*;  
RL Mol. Microbiol. 15:989-1000(1995).  
CC -|- FUNCTION: INVOLVED IN THE TRANSMISSION OF SENSORY SIGNALS FROM  
CC THE CHEMORECEPTORS TO THE FLAGELLAR MOTORS. CHEA IS  
CC AUTOPHOSPHORYLATED; IT CAN TRANSFER ITS PHOSPHATE GROUP TO EITHER  
CC CHEB OR CHEY (BY SIMILARITY).  
CC -|- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).  
CC -|- SIMILARITY: TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE  
CC KINASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U13156; AAA86674.1; -;  
DR HSP; P04002; IWF.  
DR InterPro; IPR000410; -;  
DR InterPro; IPR002545; -;  
DR InterPro; IPR002570; -;  
DR Pfam; PR01584; CheW; 1.  
DR Pfam; PR01627; Hpt; 1.  
DR Pfam; PR00512; signal; 1.  
DR PRINTS; PR00344; BCTRLSENSOR.  
KW Sensory transduction; Transferase; Kinase; Phosphorylation;  
KW Chemotaxis.  
FT MOD\_RES 46 46 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
SQ SEQUENCE 758 AA; 81110 MW; 1947225C8673831A CRC64;

Query Match 4.4%; Score 85.5; DB 1; Length 758;  
Best Local Similarity 21.1%; Pred. No. 13;  
Matches 63; Conservative 45; Mismatches 98; Indels 93; Gaps 14;  
QY 10 RNREYVSGSFANGTWTLLPKEFSAEIGPEAVTAFGLI----- 49  
Db 113 RSRLIKEALANG---ELPQ--AAASAPKTPA--GVAPAAPVVEEGFQVAFSFD 165  
QY 50 -FTTNEHIENAP----TPRGHVGSSNDPSLSYPLLIATILKLETVEVA----- 96  
Db 166 DFETGDEPTIEPTYEIVFKPSKELYAKNDATL-----LURDLSRLGEMSIHCDMT 218  
QY 97 -----AEHYGDKWNYIILT----EAMKAVIRLALFRNSGYKMLIQGETPNEEK 143  
Db 219 LPPLDRNPPEAYF---SWKISLTKDKGEAIRSVFEFAEW--DCELDVALAGTVCMD 274  
QY 144 DSNQ-----SESQNRAGNSGRNLGPHGLGNHNPWNLEGRAMSALSFFQGNAR 193  
Db 275 LPMQPPFDLSILEDAQAQAGKEDRAAASEG-----DPRNAAVAAQAQTASNVLQMAQ 327  
QY 194 TTTSTSPGNSRRITQ-----HQAQVIEPPMIK---ERRRTMSSELLTEKGVNGALFA 240  
Db 328 STARVSPENARNSSQSAQAQAQAASATPTIRVDLQKRVRLNLVGLVINVQAWLS 386

RESULT 9  
MCM3\_SCHPO  
ID MCM3\_SCHPO STANDARD; PRT; 879 AA.  
AC P30666;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE MINICHROMOSOME MAINTENANCE PROTEIN 3.  
GN MCM3 OR SPC1682.02C.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]

RP SEQUENCE FROM N.A.  
RX STRAIN=SP011;  
RX MEDLINE=98371232; PubMed=9705504;  
RA Sherman D.A.; Forsburg S.L.;  
RT 'Schizosaccharomyces pombe Mcm3p, an essential nuclear protein,  
RT associated tightly with Nda4p (Mcm3p).';  
RN Nucleic Acids Res. 26:3955-3960(1998).  
RL [2]  
RP SEQUENCE FROM N.A.  
RX STRAIN=972;  
RA Murphy L., Harris D., Lyne M., Rajandream M.A., Barrell B.G.;  
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 359-434 FROM N.A.  
RX MEDLINE=93087163; PubMed=1454522;  
RA Coxon A., Maundrell K., Kearsey S.E.;  
RT 'Fission yeast cdc21+ belongs to a family of proteins involved in an  
RT early step of chromosome replication.';  
RN Nucleic Acids Res. 20:5571-5577(1992).  
RL [1]  
CC -|- FUNCTION: PROBABLY INVOLVED IN THE INITIATION OF DNA REPLICATION.  
CC -|- SUBUNIT: ASSOCIATES STRONGLY WITH NDA4.  
CC -|- SUBCELLULAR LOCATION: NUCLEAR.  
CC -|- SIMILARITY: BELONGS TO THE MCM FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AF063864; AAC32263.1; -;  
DR EMBL; AL031525; CAA20668.1; -;  
DR EMBL; Z15034; CAA78752.1; -;  
DR PIR; S25528; S25528.  
DR PIR; S26642; S26642.  
DR InterPro; IPR001208; -;  
DR Pfam; PR00493; MCM; 1.  
DR PROSITE; PS00847; MCM\_1; 1.  
DR PROSITE; PS50051; MCM\_2; 1.  
KW Transcription regulation; DNA-binding; Nuclear protein;  
KW DNA replication; ATP-binding.  
FT DOMAIN 306 513 MCM.  
FT NP\_BIND 356 363 ATP (POTENTIAL).  
SQ SEQUENCE 879 AA; 97481 MW; 8760510C7DE2504A CRC64;

Query Match 4.4%; Score 85.5; DB 1; Length 879;  
Best Local Similarity 21.9%; Pred. No. 16;  
Matches 44; Conservative 26; Mismatches 74; Indels 57; Gaps 8;  
QY 87 KLETVEVAAEHYGDKWNYIILTAMKA---VIRLALF-----RNSGYKMLIQGG 136  
Db 633 RTLETILRLSTAH--AKARUSSVVEVKAARAEKILYALFREVVKPKRKHKKQRUEAG 690  
QY 137 -----ETPNEEKDSNQ-----SESQNRAGNSGRNL----- 161  
Db 691 EEFDSEDDNSDDMDIESEEMEDTNNVIDSGSRVTRSQNTSQSGSGSEIGSSIAGTA 750  
QY 162 GPHGLGNHNPWNLEGRAMSA-----LSFFQGNARTTSTTSGWSRRIRHQQAIVTEPP- 216  
Db 751 GSYNVGTSTNQLSWPSTHSTLPATSRLEASSDRNINTGTSVASEVSASVSEQSTVSLPRE 810  
QY 217 -----MKERRRTMSSELLTEK 232  
Db 811 KMSVFMARLASLTKSELFSEE 831

RESULT 10  
AMPM\_BACHD  
ID AMPM\_BACHD STANDARD; PRT; 248 AA.  
AC Q9Z9J4;





FT	NP_BIND	1682	1689	ATP (POTENTIAL).
FT	SITE	1773	1776	DEAH BOX.
FT	TRANSMEM	249	269	POTENTIAL.
FT	TRANSMEM	271	285	POTENTIAL.
FT	TRANSMEM	740	753	POTENTIAL.
FT	TRANSMEM	755	778	POTENTIAL.
FT	TRANSMEM	1159	1180	POTENTIAL.
FT	DISULFID	288	315	BY SIMILARITY.
FT	DISULFID	345	401	BY SIMILARITY.
FT	DISULFID	359	390	BY SIMILARITY.
FT	DISULFID	377	406	BY SIMILARITY.
FT	DISULFID	467	568	BY SIMILARITY.
FT	DISULFID	585	615	BY SIMILARITY.
FT	CARBOHYD	134	134	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	150	150	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	908	908	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	986	986	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2320	2320	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2345	2346	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2467	2467	N-LINKED (GLCNAC. .) (POTENTIAL).
SQ	SEQUENCE	3411 AA;	379524 MW;	3298C077IFED23F7 CRC64;
Query Match 4.4%; Score 84; DB 1; Length 3411;				
Best Local Similarity 18.2%; Pred. No. 1.4e+02;				
Matches 44; Conservative 30; Mismatches 58; Indels 110; Gaps 10;				
QY	104	KNWYILTEAMKAVIRLALFR	-----NSGYKMLLOG-----	135
DB	1889	KKPDFILATDAEANGANLCVERVLDCTAFKPVLDVDEGRKVAIKGLPLRISASSAAQRRGR	1948	
QY	136	-GETPNEKDS---	NOSESONRAGNSRNLGPHGLGNQHNHPW-----NLEGRAMS	183
DB	1949	IGRPNRDGDSYYSEPTSENNA-----	HHVCLWLEASMLLDNNEVYRGGM	1992
QY	184	ALSSFGQNAATTSSTPGWSRRHQOAVTEPPMIKERRRTMSELITKGVNGALFAIGE	243	
DB	1993	VAPLYGVEG-TKTPVSPG-----	EMRLDDORKVFREL-----	2024
QY	244	VLYITRLIYVLFTRKYGVRSWIPWATLSVDITGLGLANSKWGCEKSKOVHFSQPEKD	303	
DB	2025	-----VRNCDLPVLSWQVAKA-----	GLKTDNRKW-----CFEGPEEH	2058
QY	304	EL 305		
DB	2059	EI 2060		
RESULT 13				
ID	Y309_MYCPN	STANDARD;	PRT;	1325 AA.
AC	P75334;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	01-OCT-2000 (Rel. 40, Last annotation update)			
DE	HYPOTHETICAL LIPOPROTEIN MG309 HOMOLOG PRECURSOR (H08_ORF1325).			
GN	MPN445 OR MP397.			
OS	Mycoplasma pneumoniae.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;			
OC	Mycoplasmataceae; Mycoplasma.			
OX	NCBI_TaxID=2104;			
RN	SEQUENCE FROM N.A.			
RP	STRAIN-ATCC 29342 / M129;			
RX	MEDLINE=97105885; PubMed=8948633;			
RA	Himmelfreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,			
RA	Herrmann R.;			
RT	"Complete sequence analysis of the genome of the bacterium Mycoplasma			
RT	pneumoniae.";			
RL	Nucleic Acids Res. 24:4420-4449(1996).			
CC	-1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A LIPID			
CC	ANCHOR (POTENTIAL).			
CC	-1- SIMILARITY: BELONGS TO THE MG307 / MG309 / MG338 FAMILY.			
-----				
This SWISS-PROT entry is copyright. It is produced through a collaboration				
between the Swiss Institute of Bioinformatics and the EMBL outstation -				
the European Bioinformatics Institute. There are no restrictions on its				
use by non-profit institutions as long as its content is in no way				
modified and this statement is not removed. Usage by and for commercial				
entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
-----				
EMBL	AE000038;	AAE96045.1;		
DR	PROSITE; PS00013;	PROKAR_LIPOPROTEIN; 1.		
KW	Hypothetical protein; Lipoprotein; Membrane; Signal.			
FT	SIGNAL	1 27	POTENTIAL.	
FT	CHAIN	28 1325	HYPOTHETICAL LIPOPROTEIN MG309 HOMOLOG.	
FT	LIPID	28 28	N-ACYL DIGLYCERIDE (POTENTIAL).	
SQ	SEQUENCE	1325 AA;	146277 MW;	13FEF3D155ECB15 CRC64;
Query Match 4.3%; Score 83.5; DB 1; Length 1325;				
Best Local Similarity 21.3%; Pred. No. 43;				
Matches 66; Conservative 42; Mismatches 97; Indels 105; Gaps 16;				
QY	11	NREYVQSTGS-FANGLTWL-----	LPEKFSASEIGPEAVTAFGLGIPTTI	53
DB	268	NKYNOQSGGTGANGFNLLASNLKSYKNDQSKGIDIPNKFS-SDSGGKLLKASDMFDTF	326	
QY	54	NEHIENAPTGRGHVSGSGNDPSLSYPLLIALL-----	KDLETVVVEVAA--EHF	100
DB	327	-----DPFSFAAFIQGVLAQLQKSGADSKVDLSLIKDKSIIENF	366	
QY	101	YGDKKWNYIITEAMKAVIRLALFRNSG---YKMLQGETPNE-----	EKDS	145
DB	367	FVDNN-----TQA-AAAAAASSSSEGTIQIKTASDGGGTQSTVHKTDLVKIFGDKDV	419	
QY	146	NOSESONRAGNSGRN-LGPHGLGNQHNHPNLEGRAMSALSSFGQNAATTSSTPGWSR	204	
DB	420	FAGEYKQIGNTANQNTGGGSGG-----	GGGTSTGSGSTSTTTGNS----	464
QY	205	RIHQOAVIEPPMIKERRRTMSELITKGVNGA-LFAI-GEVLYITRP-----	LI	252
DB	465	----SKAVVDLTVKDKSSQPDYILSRKGDHLMAYDGGSHYLTESGRDVAQKKFLL	520	
QY	253	YVLFIRKYGCV	262	
DB	521	FRALQTKYGL	530	
RESULT 14				
ID	CP72_CATRO	STANDARD;	PRT;	524 AA.
AC	Q05047;			
DT	01-FEB-1994 (Rel. 28, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DT	01-OCT-2000 (Rel. 40, Last annotation update)			
DE	CYTCHROME P450 72A1 (EC 1.14.14.1) (CYPLXXII) (PROBABLE GERANIOL-10-			
DE	HYDROXYLASE) (GE10H).			
GN	CYP72A1 OR CYP72 OR P450CR3.			
OS	Catharanthus roseus (Rosy periwinkle) (Madagascar periwinkle).			
OC	Eukaryota; Viridiplantae; Embryophyta; tracheophyta; Spermatophyta;			
OC	Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;			
OC	Gentianales; Apocynaceae; Catharanthus.			
OX	NCBI_TaxID=4058;			
RN	SEQUENCE FROM N.A.			
RP	STRAIN-LINE CP3A;			
RA	Vetter H.-P., Mangold U., Schroeder G., Marner F.-J.,			
RA	Werck-Reichhart D., Schroeder J.;			
RT	"Molecular analysis and heterologous expression of an inducible			
RT	cytochrome P-450 protein from periwinkle (Catharanthus roseus L.)."			
RL	Plant Physiol. 100:998-1007(1992).			
RN	[2]			
RP	SEQUENCE OF 469-524 FROM N.A.			
RX	MEDLINE=93283641; PubMed=8507838;			



FT CONFLICT 838 840 RDH -> TTI (IN REF. 2.).  
SQ SEQUENCE 1142 AA; 128352 MW; 938306BC87CC5FC6 CRC64;

Query Match	4.3%;	Score 83;	DB 1;	Length 1142;
Best Local Similarity	18.7%;	Pred. No. 39;		
Matches	34;	Conservative	15;	Mismatches 65; Indels 68; Gaps 5;
QY	54	NEHIENAPTRGHVGSNDPSLSYPLLAILDKLTETVEAAAEHYGDKKNYIILTE	113	
Db	629	NPHYMNTDPKETIPYNEEDP	666	
QY	114	AMKAVIRLALFRNSGYKMLQGGTTPNEEKDSDNOSESQNRAGNSGRNLGPHGLGNQN	170	
Db	667	-----MEELSFK-----EDPTVRHYEGEQTYSNQPKELYPLSLDNPSPKPR	706	
QY	171	-----HHNPWN-----LEGRAMSALSFFQGNARTTTSTPGWSRRIQ	207	
Db	707	EDFLYLGFEYPNPEENFPSTYNTAPTVSSPVESSRGYANNAGQEESTMPFSSWSDPRIQ	766	
QY	208	HQ	209	
Db	767	QA	768	

Search completed: July 11, 2001, 11:36:17  
Job time: 76 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2001, 00:15:10 ; Search time 2011.93 Seconds  
(without alignments)  
11401.327 Million cell updates/sec

Title: US-09-545-072A-1  
Perfect score: 1483  
Sequence: 1 attgcaccagagagagaaaa.....attaaaaaaaaaaaaaaaaaaaaa 1483

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues 2688314

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*

1: gb\_ba1:\*

2: gb\_ba2:\*

3: gb\_ba3:\*

4: gb\_in1:\*

5: gb\_in2:\*

6: gb\_in3:\*

7: gb\_om:\*

8: gb\_ov:\*

9: gb\_pat1:\*

10: gb\_pat2:\*

11: gb\_ph:\*

12: gb\_pl1:\*

13: gb\_pl2:\*

14: gb\_pl3:\*

15: gb\_pl4:\*

16: em\_ba1:\*

17: em\_ba2:\*

18: em\_fun:\*

19: em\_htgo\_hum:\*

20: em\_htgo\_inv:\*

21: em\_htgo\_rod:\*

22: em\_htg\_hum1:\*

23: em\_htg\_hum2:\*

24: em\_htg\_hum3:\*

25: em\_htg\_hum4:\*

26: em\_htg\_hum5:\*

27: em\_htg\_hum6:\*

28: em\_htg\_hum7:\*

29: em\_htg\_hum8:\*

30: em\_htg\_inv1:\*

31: em\_htg\_inv2:\*

32: em\_htg\_other:\*

33: em\_htg\_rod:\*

34: em\_hum1:\*

35: em\_hum2:\*

36: em\_hum3:\*

37: em\_hum4:\*

38: em\_hum5:\*

39: em\_hum6:\*

40: em\_hum7:\*

41: em\_in:\*

42: em\_om:\*

43: em\_or:\*

44: em\_ov:\*

45: em\_pat:\*

46: em\_ph:\*

47: em\_pl:\*

48: em\_ro:\*

49: em\_sts:\*

50: em\_sy:\*

51: em\_un:\*

52: em\_vi:\*

53: gb\_sts1:\*

54: gb\_sts2:\*

55: gb\_sts3:\*

56: gb\_sy:\*

57: gb\_un:\*

58: gb\_vil:\*

59: gb\_v12:\*

60: gb\_htg1:\*

61: gb\_htg2:\*

62: gb\_htg3:\*

63: gb\_htg4:\*

64: gb\_htg5:\*

65: gb\_htg6:\*

66: gb\_htg7:\*

67: gb\_htg8:\*

68: gb\_htg9:\*

69: gb\_htg10:\*

70: gb\_htg11:\*

71: gb\_htg12:\*

72: gb\_htg13:\*

73: gb\_htg14:\*

74: gb\_htg15:\*

75: gb\_htg16:\*

76: gb\_htg17:\*

77: gb\_htg18:\*

78: gb\_htg19:\*

79: gb\_htg20:\*

80: gb\_htg21:\*

81: gb\_htg22:\*

82: gb\_htg23:\*

83: gb\_htg24:\*

84: gb\_htg25:\*

85: gb\_pri:\*

86: gb\_pr2:\*

87: gb\_pr3:\*

88: gb\_pr4:\*

89: gb\_pr5:\*

90: gb\_pr6:\*

91: gb\_pr7:\*

92: gb\_pr8:\*

93: gb\_pr9:\*

94: gb\_rol:\*

95: gb\_rod:\*

96: gb\_in4:\*

97: gb\_pri0:\*

98: em\_ba3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1483	100.0	1483	12	AF085354 Arabidops
c 2	739.6	49.9	89794	12	AC004665 Arabidops
c 3	48.6	3.3	318221	84	PFMAL13P3 AL049184 Plasmodiu
c 4	45.6	3.1	3791	6	D50338 Dictyosteli
c 5	44.2	3.0	1141	10	AX083744 Sequence
c 6	44.2	3.0	3747	97	U42766 Human neuro
c 7	44.2	3.0	140658	8	AF112374 Danio rer
c 8	44.2	3.0	183511	62	AC012498 Homo sapi





Qy 721 cgggttgctcgaagaattcaacatcagcagcagttatagagctccaatgatcaagga 780  
 |||||  
 Db 721 CGGTGGTCTCGAAGAAATCAACATCAGCAGCAGTTFATAGAGCTCCAATGATCAAGGA 780  
 Qy 781 gagcggaagaacgatgtccgagctacttacttgagaaaggttgaatgagcgtgtttgc 840  
 |||||  
 Db 781 GAGCGGAAGAACGATGTCGAGCTACTTACTTGAGAAGGCTGTAATGAGCGTGTGTTGC 840  
 Qy 841 gattggtgagttctttacataaagagagacccgctcatcttcaatctttcatcagaata 900  
 |||||  
 Db 841 GATTGGTGAGGTCTTTACATAAGCAGACCCGCTCATTTACGTCTCTTTTCATCAGAAAAATA 900  
 Qy 901 tggagtcgagcttggtatctctggcttatatcgctttcttgagacacactgggagtgag 960  
 |||||  
 Db 901 TGGAGTCCGATCTGGATTCCTTGGCTATATCGCTTCTCTGGACACACTGGGATGGG 960  
 Qy 961 tcttttgcgaattcgaagtggtggtgagagagagcaagcaagtcattcttcaggagcc 1020  
 |||||  
 Db 961 TCCTTCTTCCAAATTCGAAGTGTGGGGAGAGCAAGCAAGTCCATTTCTCAGAGACC 1020  
 Qy 1021 tgaagaagatgagctgagagacacaaaactgattgagcattgtacctcatgagagatcc 1080  
 |||||  
 Db 1021 TGAAGAAGATGAGCTGAGGAGACGAAACTGATATGGGCATTTGACCTCATGAGAGATCC 1080  
 Qy 1081 attcttcaccaagtacacaaagcagaagctggaagctctcaaaagagcgtggaactaat 1140  
 |||||  
 Db 1081 ATTCTTCCACCAAGTACACAAAGCAGCAAGCTGGAAGCTCTCAAAAGAGCTGGAACATA 1140  
 Qy 1141 tccattgctgagttcctcacagagaaagtgtgagctgtttgagggagctcagtcacg 1200  
 |||||  
 Db 1141 TCCATTGATGAGATTCCTCACAGAGAAGATTGTGGAGCTTTTGGAGGAGCTCAGTCAGC 1200  
 Qy 1201 gtacacttacatctcgatcgtgaagtgtaagcgttttacttatgttttatgcaacgg 1260  
 |||||  
 Db 1201 GTACACTTACATATCGGATCGTAGGTTAAGCGTTTACTATGTTGTTATATGCAACGG 1260  
 Qy 1261 aagaataattgcattgttggagtgcttttttagatcatcaagagcctcctacagattctt 1320  
 |||||  
 Db 1261 AAGATATTGCAATGTTGGAATGCTTTTGTAGATCATCAAGGCTCCTCAGATTTCTT 1320  
 Qy 1321 aggaatggtttcagcgtttgttagaattgtgtttatttcaacaggtagagacataa 1380  
 |||||  
 Db 1321 AGGGAATGGTTTCAGGCTTTTGTAGAATTGTGTTTATTCACACAGGTAGAGAACATAA 1380  
 Qy 1381 ccatagacagatgtatctgaagagataagctctctatgtcttaagaagaatggaccgatac 1440  
 |||||  
 Db 1381 CCATAGACAGATGATCTGAGAGATAGCTTCTCTATGTCTTAACAATGGACCCATAC 1440  
 Qy 1441 gaataaaacagcatcattaaagattaaataaaaaa 1483  
 |||||  
 Db 1441 GAATAAAACAGCATCAATTAAGATTAAAAA 1483

RESULT 2  
 LOCUS AC004665 89794 bp DNA PLN 05-APR-2000  
 DEFINITION Arabidopsis thaliana chromosome II section 244 of 255 of the complete sequence. Sequence from clones F17K2, F4118.  
 ACCESSION AC004665 AE002093  
 VERSION AC004665.2 GI:6598428  
 KEYWORDS HTG.  
 SOURCE thale cress.  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids  
 1 (bases 1 to 89794)  
 Lin, X., Kaul, S., Rounsley, S.D., Shea, T.P., Benito, M.-I., Town, C.D., Fujii, C.Y., Mason, T.M., Bowman, C.L., Barnstead, M.E., Feldblyum, T.V., Buell, C.R., Ketchum, K.A., Lee, J.J., Ronning, C.M., Koo, H., Moffat, K.S., Cronin, L.A., Shen, M., VanAken, S.E., Umayam, L., Tallon, L.J., Gill, J.E., Adams, M.D., Carrera, A.J., Creasy, T.H., Goodman, H.M., Somerville, C.R., Copenhaver, G.P., Preuss, D.,

NIERMAN, W.C., WHITE, O., EISEN, J.A., SALZBERG, S.L., FRASER, C.M. and VENTER, J.C.  
 Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana  
 Nature 402 (6763), 761-768 (1999)  
 20083487  
 10617197  
 2 (bases 1 to 89794)  
 Lin, X.  
 Direct Submission  
 Submitted (09-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA  
 On Dec 17, 1999 this sequence of chromosome 2 was merged from those of the individual clones on this chromosome after removing overlaps. For detailed information, please see the TIGR web site (http://www.tigr.org/tdb/at/at.html).

Genes were identified by a combination of three methods: Gene prediction programs including GRAIL (ftp://arthur.epm.ornl.gov/pub/xgrail), Genefinder (Phil Green, University of Washington), Genscan (Chris Burge, http://gnomic.stanford.edu/GENSCAN.html), and NetPlantGene (http://www.cbs.dtu.dk/services/NetGene2/), searches of the complete sequence against a peptide database and plant EST databases at TIGR, and manual curations based on those analyses. Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by two or more gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats were identified by RepeatMasker (Arian Smit, http://ftp.genome.washington.edu/RM/RepeatMasker.html). Genes are numbered from the top to bottom of the chromosome.

We thank the CSHL/WashU/ABI consortium for sequencing BAC clones F6223, F536, T17A5, and T13L16, the ESSA group for sequencing clone F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards and Satoshi Tabata for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics Department, especially Lixin Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy Peterson, Michael Holmes, and Delwood Richardson for software and database support.

This work was supported by the National Science Foundation, Department of Energy and the US Department of Agriculture.

Address all correspondence to: at@tigr.org.

FEATURES  
 source  
 1..89794  
 /organism="Arabidopsis thaliana"  
 /cultivar="Columbia"  
 /db\_xref="taxon:3702"  
 /chromosome="II"  
 misc\_feature  
 <1..2743  
 /note="Sequence from clone F17K2"  
 complement(join(<940..1001,1102..1163,1345..1412,1457..2051,2136..2355,2646..>2721))  
 /gene="At2g45690"  
 complement(<940..>2721)  
 /gene="At2g45690"  
 /note="F17K2.22; predicted by genscan and genefinder"  
 complement(join(940..1001,1102..1163,1345..1412,1457..2051,2136..2355,2646..>2721))  
 /gene="At2g45690"  
 /note="hypothetical protein"  
 /codon\_start=1  
 /protein\_id="AAC28524.2"  
 /db\_xref="GI:6598429"



Db	799	AGGTAGAACATTAACCATAGACAGATGATCTCTGAAGAGATAAGCTTCTCTATGCTAAA	740
Qy	1426	gaatgacccatcacgaataaaacaagcatcattaaagtataa	1469
Db	739	GAAATGGACCATTAGCAATAAACAAGCAYCATTAAAGATTAAA	696
RESULT	3		
PFMAL13P3/c			
LOCUS	PFMAL13P3	318221 bp	DNA HTG 19-AUG-1999
DEFINITION	Plasmodium falciparum chromosome 13 strain 3D7, *** SEQUENCING IN PROGRESS ***, In unordered pieces.		
ACCESSION	AL049184		
VERSION	AL049184.5	GI:5763803	
KEYWORDS	HTG; HTGS, PHASEL.		
SOURCE	malaria parasite P. falciparum.		
ORGANISM	Plasmodium falciparum		
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
AUTHORS	Bowman,S., Churcher,C., Harris,B., Harris,D., Lawson,D., Quail,M. and Barrell,B.		
TITLE	Direct Submission		
JOURNAL	Submitted (15-MAR-1999) P.falciparum Genome Sequencing Consortium, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK		
COMMENT	On Aug 24, 1999 this sequence version replaced gi:5731882. For more information about this sequence or the Malaria Project, see http://www.sanger.ac.uk/projects/P_falciparum. IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc. Order of segments is not known; 800 n's separate segments. * NOTE: This is a 'working draft' sequence. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved.		
FEATURES			
source	Location/Qualifiers		
	1..318221	/organism="Plasmodium falciparum"	
		/strain="3D7"	
		/db_xref="taxon:5833"	
		/chromosome="13"	
BASE COUNT	113610 a 29264 c 27321 g 118405 t 29621 others		
ORIGIN			
Query Match	3.3%; Score 48.6; DB 84;	Length 318221;	
Best Local Similarity	64.9%; Pred. No. 0.14;		
Matches	72; Conservative 0; Mismatches 39; Indels 0; Gaps 0;		
Qy	1373	gaacataaccatgacagatgtatctgaagagataaagttcttctatgtcttaagaagattgg	1432
Db	12715	GAATAAAAAATATATATATATTTATATATTTATGCATTATATTTTTATTAAAAAGA	12656
Qy	1433	accgatcgataaaaacaagcatcattaaaggataaaaaa	1483
Db	12655	AACGACACNAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	12605
RESULT	4		
D50338/c			
LOCUS	D50338	3791 bp	DNA INV 14-APR-2000
DEFINITION	Dictyostelium discoideum DNA for ribosomal protein, complete cds.		
ACCESSION	D50338		
VERSION	D50338.1	GI:790953	
KEYWORDS	ribosomal protein.		
SOURCE	Dictyostelium discoideum (strain:AX-3) vegetative growth amoeba cell DNA, clone.lib:HincII limited lambda gt10 clone:lamba GH1.		
ORGANISM	Dictyostelium discoideum		





```

/product="odorant receptor 5.3"
/protein_id="AAF27260.1"
/db_xref="GI:6739473"
/translation="MGSRNVEYQNTTIITEFVLGRGPPGLPPEVHSGVALLFYLIL
VLATGNIFIVMEIKSQPTVIIFCNLAMDALGALGTTTPRVIAKYMWDKILISF
SCQTFQMFVFLGATNLSFALMALDRFAICNPLRYHTLTKNLTILICAFWTAN
MLOLVGVLLSVKSCSNVPHCYCDHVSISKLACDAATMKVTSTAIAMFWLWGP
LSFILSVSIIISVSKISGEGRYKFTSTCPOLLIIICLYLPRTFYIITNISGYEL
STDTRVMVMYSLIPAVINPFYCFYCFYCFYCFYCFYCFYCFYCFYCFYCFYCFY
complement(<20838..>21887)
/gene="OR 5.1"
/product="odorant receptor 5.1"
/complement(<20838..>21887)
/gene="OR 5.1"
/complement(20838..21887)
/gene="OR 5.1"
/codon_start=1
/product="odorant receptor 5.1"
/protein_id="AAF27261.1"
/db_xref="GI:6739474"
/translation="MYCTLKSACVNVNNAISILVASVEKLGSSVMMAYTNITE
FYLLGPGHPEYGAAGFLFLVLTLAGNIFIIISYAEKSLQKPTLYLFCNLAV
CDIATGVTLPKAIKAYFEDSSEITFLSCFQMFVHYLGSINFLMLLMAVDRFVAV
CNPLRCVLIYNTKTLTLLACALIMLCISMLTVMYNTFDEPYCGSNITONCYDHSI
IKLSGDIQKQKFFSFAFVLLGGLFIIFSETAIIISVFLSDTQARPKTSTCT
POLLICILYPRCVYVYVHDVIAISPGIRVNLINWYILPIIPVNPMLICFTRKID
AIKKLKKDKVNVQI"
/repeat_region
25843..25887
/rpt_type=tandem
/rpt_unit=ca
/complement(27763..28719)
/gene="OR 13.4p"
/note="odorant receptor 13.4"
/pseudo
/complement(31126..32258)
/gene="OR 13.5p"
/note="similar to OR 13"
/pseudo
/complement(<34090..>35034)
/gene="OR 13.2"
/product="odorant receptor 13.2"
/complement(<34090..>35034)
/gene="OR 13.2"
/complement(34090..35034)
/gene="OR 13.2"
/codon_start=1
/product="odorant receptor 13.2"
/protein_id="AAF27262.1"
/db_xref="GI:6739475"
/translation="MSVGNSTFAKDFAIVGPPGLQPHYGLVSLLEFVYVLTILGNA
VFFTLFVITKSLRPVYFIINLVASOVLFSTTLTKIISRYFQDGTISFLGCFVOM
FVHFLGSACSVLVAIMADRYTAICYPLQYHSIMNKNVLLILSPWILSFWGFLAL
VIRAYPLPYCADNSIIHCYCDHVSITTLACTDRSLYSIPALIYAFLLILVPPFAVIFS
YCSIFIAVMRISGSGRMKFTSTCPOLLIIIALYFIPRVNYLSSNIGLKAFSTDLRII
MIMYSVLPLINPIIYCLRTEEVKVKLRQHFSEKTR"
/repeat_region
35978..36129
/rpt_type=tandem
/rpt_unit=gaga:gata
/complement(<39727..>40668)
/gene="OR 13.3"
/product="odorant receptor 13.3"
/complement(<39727..>40668)
/gene="OR 13.3"
/complement(39727..40668)
/gene="OR 13.3"
/codon_start=1
/product="odorant receptor 13.3"
/protein_id="AAF27263.1"
/db_xref="GI:6739476"
/translation="MPVENTSFVKDFVIVGPPGLQPHYGLVSLLEFVYVLTILGNA
VFFTLFVITKSLRPVYFIINLVASOVLFSTTLTKIISRYFQDGTISFLGCFVOM
FSVHLFGVSCSVLVAIMADRYTAICYPLQYHSIMNKNVLLILSPWILSFWGFLAL
VIRAYPLPYCADNSIIHCYCDHVSITTLACTDRSLYSIPALIYAFLLILVPPFAVIFS

```

```

YCSIFIAVMRISGSGRMKFTSTCPOLLIIIALYFIPRVNYLSSNIGLKAFSTDLRII
MIMYSVLPLINPIIYCLRTEEVKVKLRQHFSEKTR"
/complement(<42101..>43087)
/gene="OR 13.1"
/product="odorant receptor 13.1"
/complement(<42101..>43087)
/gene="OR 13.1"
/complement(42101..43087)
/gene="OR 13.1"
/codon_start=1
/product="odorant receptor 13.1"
/protein_id="AAF27264.1"
/db_xref="GI:6739477"
/translation="MSAGNISFYKFEFIVGPGKQGYFDSVAALMLFVFCILAGNG
TFLILFSLKSLHMPYVILNLVADILFSTTLTKIARYLFDGGISFMCQFIOM
YFVHFGTVNAFILAIFADRYVAITCNPLRPVNTKWNIFYLCLISWIANVTPLMW
VIRAYPLPYCAENTIIQCYCDHISITSLACTNRAPYSIPAFVLAWALLTPLAIFVES
XCALILAVLRISSTQARLKTFTSCPOLIIIALYFIPRCFIYLSNIGIYFFSTDRLA
IIMYSLFFPMFNPFLIYCLRTKDKVCEKLVARLKAISTNVKFNLTKVDPISVSS"
46118..46169
/rpt_type=tandem
/rpt_unit=ta
/48958..48958
/rpt_type=tandem
/rpt_unit=ca
/50334..51311
/gene="OR 2.8"
/product="odorant receptor 2.8"
/50334..51311
/gene="OR 2.8"
/50334..51311
/gene="OR 2.8"
/codon_start=1
/product="odorant receptor 2.8"
/protein_id="AAF27265.1"
/db_xref="GI:6739478"
/translation="MSDSASFSSNISIVHPYEFFITGLSGTEYSNYVYIFLFFLIIV
AVIGNFVLLIILALERSHPKYIAVFNALADIGEMNALVPMNKMLFESQYISIN
ACVMNMFVHFFHTMQSTFLVMSYDRFAICLPLRYHAIVNNTVISVFLVLMFNV
FVLLVYVGMITRLSICKSNVLSYCDHGVPKACNDNSINFLGKLSLALYLVAPL
IIIFLSYGVFLALSKITWEGRLKALKTCVAHLLLAGSFFVPIFISYIATSMNSLTP
NARVLSLSVYLPPLTNPILYLVNTAEIKLKIKNRSTKTVHVAS"
<60772..>61752
/gene="OR 2.11"
/product="odorant receptor 2.11"
/60772..61752
/gene="OR 2.11"
/60772..61752
/gene="OR 2.11"
/codon_start=1
/product="odorant receptor 2.11"
/protein_id="AAF27266.1"
/db_xref="GI:6739479"
Query Match 3.08; Score 44.2; DB 8; Length 140658;
Best Local Similarity 49.08; Pred. No. 1.7;
Matches 118; Conservative 0; Mismatches 123; Indels 0; Gaps 0;
Qy 1243 atggtttatatgcaacggaagaattgtccattgttggaattgttttagatcatacaaa 1302
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 119779 ATGCTTGATATGCCAAAATATTATATACATTTATTATTATTATTATTATTATT 119720
Qy 1303 ggtcctcacagattcttagggaatggttcaggcttttttagaagaattgctttattgc 1362
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 119719 TAATCATATATTTATGATTTGAAATTTTATGACAAATCTTATTGTCATATCTGTTT 119660
Qy 1363 aacaggtagagaacataaacatagacagatgatatcgaagagataagctctctatgct 1422
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 119659 TAATGTTTCTAATATTTTATTAAGTGTCTAAGCAGAACACGCTAGTGTAAATGT 119600
Qy 1423 aagaagaatgacgcacagataaaaaaacagcatcattaaagattaaaaaataaaaaa 1482
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 119599 GCCTATTTTAGGTGTAAATATACCAAGAAAACAAAATGGCAAAAAAATAAAAAA 119540

```











KEYWORDS HTG; rifin; telomere; var; var-like hypothetical protein.  
SOURCE malaria parasite P. falciparum.  
ORGANISM Plasmodium falciparum  
REFERENCE Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
AUTHORS Oliver, K., Bowman, S., Churcher, C., Harris, B., Harris, D., Lawson, D., Quail, M., Rajandream, M. and Barrell, B.  
TITLE Direct Submission  
JOURNAL Submitted (24-SEP-1998) P-falciparum Genome Sequencing Consortium, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK  
COMMENT On Apr 28, 2000 this sequence version replaced gi:5731895.  
For more information about this sequence or the Malaria Project, see [http://www.sanger.ac.uk/Projects/P\\_falciparum](http://www.sanger.ac.uk/Projects/P_falciparum).  
FEATURES  
source  
1..66441  
/organism="Plasmodium falciparum"  
/strain="3D7"  
/db\_xref="taxon:5833"  
/chromosome="1"  
repeat\_region  
1..579  
/note="telomeric repeats"  
join(29733..34985,36111..37349)  
/gene="var"  
/note="MALIP4.01"  
join(29733..34985,36111..37349)  
/gene="var"  
/note="MALIP4.01, var gene, len: 2163 aa, similarity: eg to Q25733 PFEMP1 variant 1 of strain MC (2924 aa) fasta scores: opt: 3681, E(): 0, 35.8% identity in 3011 aa overlap"  
/codon\_start=1  
/product="var, MALIP4.01"  
/protein\_id="CAB89209.1"  
/db\_xref="GI:7670005"  
translation="MYTSSGGGGAAGSGEEDAKHVLDEFQGVYNEKVEYKSIY KEALGDSLSAGTSCALEYEEYKHTNGGKGKRYCTELGEKVEPRFSD KALGDSLSAGTSCALEYEEYKHTNGGKGKRYCTELGEKVEPRFSD VCGAAYEGDSIKYNYPKYQYPTDNLQCTVLARSFADIGDIVRGDLYLGNLQVES TAITLLENNKIDIFAKIHSVDYSGSNGALQKRYKOTDNYELREDWALNDDQVW KAITCNAGGNGYFQTCGSGEMWCKCRCKDKVPTDYVPOYLRFWEFADDFCR LRKHLKANKRCGSDNDRYCDLNRDCTOTIRNGHEHFEKDKCKCQVSCAHFVN WIDNOKLEPKEREKTYEIKKKHPTTIIITANKRKTINILNYKFEYKYLQEKYGDV ENFLKNEQECQNPYNDESSIDINFKSIDIDIFSHTEYCAQPCWCAKRGKGW EPEKCTGKTYDPEKKTNPILPTYSIQSILKKNKFCNGNGGAGAPATATGCGG IKWCHYEGDNDNCVEGEWEKFEKGNVSNYAFWKKVHMDLIDSMOWATNHEGNC INKNDNTCKNSCKRPECFCFHWQKKNEWEAKDHEKONTAAETCCDQGVYLOW VILDFLDRESDENKVSABEAEIKHLQMLQOAGVDPAPAFRCPTEDGVAEQD TIMDLLNRENDATECKCDKPPAPPTAGPGARADPHDVQPPPGSGGTGAN DEEDDDDDDDDEDEGEAKEEDEEEDQEDVHQEAKKEPEKEEVARTPKDYNVC NIVNVTDSGSLQAACSLKYGNAPTSMKVCYTPSGNTSDTVKSGDTGSGICVPPR RRLXYPLTRLTGSDSTQASQASEVOTQARSGNTDKSPGSEAAQGDGVSQDPKA LKAFVESAATFELWDRYKKIKEKEKKEKTYEQIYESTDYDEEKDQOEELKKG IIPDEFKRMFTYLDGYDLYLSDGTVANGNEKILKAINNYFOKIREQSSDNNPSP RSVKPTSDKQPTQWNAHAPSWNAWCALITYDNGSGGKKTITITQDNLKIALW DENGKPLKTYQDSVTIGASGAKPQAKPTGDTPLTQVLFVRLFRYLEWGQN FCKRTMLEKIKYCKYQGGGRKQTPQCSYGENCDQDLDNDSTADLCKPGC GRCERYKWKIEKKEEFTKQSNVEQKTKQCKESAKGNHNGNEFCQGTCDTA GDFNLKSGPCKKENGKNOEDENFEDKEDTFCGHENVCAPVFKDICKKDCORNA SNMCKGDFITAEADIKIMDSSEVNMVSDNDPNFKPDGLDACKDAHIFPKIKENK WSCNGVCGTNPCKRYNGEKGSGENNOIITIRGLVTHVQNFLLDNNKIKTKLKPC RANGESVKICKDVKKKEKTEEPKIRDRYLEPYKSDGYNKSLVRSMEETIPL MDTNGKERIQELNFKLSEYECNADNSQGDTPKIDVECLLEKEDKANKKCTQTS GTDCHPSTPLEDDPELEETENPCTKQPEKEDGCEAAETAEETSPTA TSEGTEQNSPPPPAPAPAPAPAEQSPKQKVEPQPPQPTNPPNPNFNNPVI QKTDHPDYSKGDIPLNTQNTLYFNKPEEPFTISIDRLDLYTGEQISYINHTSNT MDPKYVSNNYSGDILNDALNGDYDIYDEILKRKENELFTGNHVKQTSIHSVAKLT NSDPTHNOLFEKHLDRNRNCEKWNDRNLAKEWENETHSGNTSDSNKTL NDVDSIQIDMHEKRMKEFTNMDDTLDDLKTYNEPYDVQDDIYYDVNDHDASTVDSN NMDVPSRVQIEMDVNTKLVEKYPADWADI"

34986..34991  
/note="potential splice donor sequence, aag/gtatta"  
36101..36110  
/note="potential splice acceptor sequence"  
complement(join(39205..40146,40377..40430))  
/gene="rifin"  
/note="MALIP4.02"  
complement(join(39205..40146,40377..40430))  
/gene="rifin"  
/note="MALIP4.02, rifin, len: 331 aa, similarity: most to 097325 PFC0010C protein (326 aa), fasta scores: opt: 1526, E(): 0, 72.2% identity in 331 aa overlap"  
/codon\_start=1  
/product="rifin, MALIP4.02"  
/protein\_id="CAB89210.1"  
/db\_xref="GI:7670006"  
translation="MKIHYINILFELPILNLIYNQNHKSTPHTPNHTQTRLLCE CELSPANNDAMKRVMOQFEDRTQRFHEYDERMKTTRQCKQCKOKEIQIILK KLEKELMDKRFATLQTDIQSDISPTICEKSLKEDKVEKGLRCAGVLGGIAPGWSLV SGLGYAVTNYVTQALQIEAGVKGAGIELGRDFSLGKGLIPISVQLNLIHNTYDI AKTYITPKSVNSTKCAVKHSHFCSTYISNENALSKRAAGIAEYAADMAKITERGVL DAATPGLTYSNAITASVAIVLVMIILYLLRYRRKKMKKKLQYIKLLEE"  
complement(40147..40156)  
/note="potential splice acceptor sequence"  
complement(40371..40376)  
/note="potential splice donor sequence, ttg/gtaatt"  
complement(join(42589..43839,43997..46729))  
/gene="MALIP4.03"  
complement(join(42589..43839,43997..46729))  
/gene="MALIP4.03"  
/note="MALIP4.03, var-like hypothetical protein, len: 1327 aa, similarity: most to 96294 PFEMP1 fragment (431 aa), fasta scores: opt: 1852, E(): 0, 67.0% identity in 440 aa overlap"  
/codon\_start=1  
/product="var-like protein, MALIP4.03"  
/protein\_id="CAB89211.1"  
/db\_xref="GI:7670007"  
translation="MALKGVINEKSLARNVLENTANEIKVRENERKYNKDLKGLS KEQSDGSLSSGDIRSGYDSCSLDHKPHNTINNGYPPARNPCDGNQERFSNDGE SKGSDKIRGNENNSNAGACAPFRONCKNLEYLINKNTENTHOLLGNLVTAKEYE GESVNNHPDKLKGKSSICTSLARSFADIGDIIRGDMFKRKHNDNIKGVLYRVEFK KITEGLNKGARHEYKVGKNGYIKLREDWNTANROQVAKIACEAPENAYILKRID GDGIDNLIITHPKCGHDTDPVVDYIPQRLWMSSEYFCNVLENKEIDEMNQCKDC ESAPERYRRNRHRRGDDDTNVQLFLKYIENNEKQESLGKLYDKTSCHTGNFNY DNTPGSRNFAETTPKFKKACKIPNLEKPCPENENKNDKVDKVSCHLSLSPK NDLSEWNSGVKNKENDNGVLPVPRRLCINLFSKKDYKMKDNDKEDLLNAAFS QGKLLGKKYSYNEAEYKFSYADYSDIVKGTDMNDLKLNLKELTLLKTEKGD ISVDRKTWDDNDKRVNWMALCGTKTENQQLNQLDITVECKRLFDIYEEFRYRNQWGL QYCKELIKAHINTKCDIEVGRKHKSMVDITDVECKRLFDIYEEFRYRNQWGL SERIKITKSKNSQVNIPECAASYVTKHNGCINLRDMDIHKNNINNNELKEM INITKFTQDQYRQLOQINSMEINTKSVKTAVDTTKDIVSYGLAGTMGVAALQAG DFLGKTDQLYNEMFKPVEKKLDTSSKNLNIYEDPNIMVAGIYVALTLGLLFLKWT KPRHEVDRILOMSONKYCIPTKSPNKYIPVASORYKGTIYVWEGSDSDKDYIG NISSDDTISESETEEMDINDIPTYSPKTKLIEVLEPSKNGTPTSGKDNFTVD DMVPTTDFDESELSKHDIFSNMLQNDPNVDKSGDIPNTQPTNLTLYFDNQOE KPFTTSTHDDLYTGEESYINHTSMDPKYVSNNYSGDILNDLNTLSHQHDI KEDILKRNKELFNTNPKNTSNNSVAKLTNSDPIIMQLDLHKDLDRHRCMDKWT YDELLKLNQWNNKNNVGGDIPNDKMLNTDSIQIDIDENKGLDKLSDNMDTILDDM EDDIYDVNDNDNDQDPSYDVISMDHNDYDVDPVKVHTEMILNNTSNGSLEQQFP ISDVWNI"  
complement(43840..43849)  
/note="potential splice acceptor sequence"  
complement(43991..43996)  
/note="potential splice donor sequence, aag/gtaatt"  
join(50585..50638,50794..51858)  
/gene="rifin"  
/note="MALIP4.04"  
join(50585..50638,50794..51858)  
/gene="rifin"  
/note="MALIP4.04, rifin, len: 372 aa, similarity: to rifins, most to 097327 PFC0035W protein (374 aa), fasta

```
scores: Opt: 1216, E(): 0, 55.9% identity in 374 aa overlap"
/codon_start=1
/product="rifin, MALLP4.04"
/protein_id="CAB89212.1"
/db_xref="GI:7670008"
/translation="MKIHYTNILLFPKLNLVNMHKPSITSRRHTQTTRLLCEBELY
TNYDNDPEKSMYQOFDRDTORFHEYDENLKEKRQCKDKRIQKIILKDRIE
ETEFKSQITDHSDAIPTCIKESLADKVKNCLCTQNLIKIVAPSGVLAGEIS
AALSVWKTTEIAAMAEKLAGAGAKGLHAGTAGNNAVIEQLRTLGIFYFGVKLLLETI
IDVTNNYSFIYDKVYSHVTSCTPSLVNDOLAVGTENTSDPCNLVHNSLOGSFYRS
SQATIIKVEEAQAEGAANTTKVAIMPVIYTTFEAKNAIEAVEAATSYYPPIIAS
IVAIVIIMVIIYLIRRYRKKMUKKKLYIKLEE"
/misc_feature
50639..50644
/note=potential splice donor sequence, ttg/gtaaca"
/misc_feature
50784..50793
/note=potential splice acceptor sequence"
/misc_feature
complement(53425..53430)
/note=potential splice donor sequence, aag/gtaaata, at 3'
end of MALLP4."
gene
complement(53431..53502)
/genome="MALLP4.05, partial"
/complement(53431..53502)
/partial
/genome="MALLP4.05, partial"
/note="MALLP4.05, partial, len: 24 aa, potential
exon/pseudogene, similarity: PFBI045W, O96294 PFEMP1
fragment(431 aa) fasta score: 95.8% identity in 24 aa
overlap"
/codon_start=1
/product="potential exon/pseudogene, MALLP4.05"
/protein_id="CAB89213.1"
/db_xref="GI:7670009"
/translation="MMSSIAFPSLVGIFAFATLSFYFLK"
/complement(54000..57115)
/genome="MALLP4.06"
/note="rifin"
complement(join(54000..55010,55160..55228))
/genome="rifin"
/note="MALLP4.06, rifin, len: 359 aa, similarity: most to
O97326 PFC0030C protein (359 aa), fasta scores: opt: 1549,
E(): 0, 68.5% identity in 359 aa overlap"
/codon_start=1
/product="rifin, MALLP4.06"
/protein_id="CAB89214.1"
/db_xref="GI:7670010"
/translation="MKIHYTKLLFFLTNILTSYAHNKPKPTSPPHHRTSTSRLL
LSVDYTPESSTDDEIDSVEKFEPQASORLEYDERLERQEQRKEQPDKNIOKLI
HKDMENKLAERKETECRCGGCSGVAGSIGDFGAVINWRPALDAITAAILANLN
AVKIAAAAANAAGEVGKALVIBELQMGISTLKQTLKSPFTISYNKVVSSITKAYIG
```

repeat_region	/note="MER45 repeat: matches 1. .178 of consensus" 5673. .5759
repeat_region	/note="L2 repeat: matches 2655. .2744 of consensus" 5760. .6044
repeat_region	/note="AluSx repeat: matches 9. .294 of consensus" 6045. .6280
repeat_region	/note="L2 repeat: matches 2424. .2655 of consensus" 6494. .6793
repeat_region	/note="MER33 repeat: matches 11. .311 of consensus" 7461. .7550
misc_feature	/note="45 copies 2 mer tt 63% conserved" 8007. .8010
repeat_region	/note="IS5 excised. Annotated bases relate to duplicated flanking sequence at point of insertion." 8775. .9103
repeat_region	/note="MLTIB repeat: matches 1. .390 of consensus" 9258. .9567
repeat_region	/note="AluSc repeat: matches 5. .307 of consensus" 9740. .9777
repeat_region	/note="19 copies 2 mer tg 84% conserved" 10555. .10831
repeat_region	/note="Aluwb repeat: matches 1. .273 of consensus" 10952. .11103
repeat_region	/note="L1MB4 repeat: matches 6026. .6185 of consensus" 11597. .11763
repeat_region	/note="L1MC3 repeat: matches 7569. .7739 of consensus" 11764. .12073
repeat_region	/note="AluSx repeat: matches 1. .310 of consensus" 12074. .12152
repeat_region	/note="L1MC3 repeat: matches 7490. .7569 of consensus" 12153. .12462
repeat_region	/note="AluSx repeat: matches 1. .311 of consensus" 12463. .12888
repeat_region	/note="L1MC3 repeat: matches 7083. .7490 of consensus" 12897. .13482
repeat_region	/note="L1MC3 repeat: matches 6142. .6651 of consensus" 13483. .13788
repeat_region	/note="AluSx repeat: matches 1. .306 of consensus" 13789. .14189
repeat_region	/note="L1MC3 repeat: matches 6651. .7073 of consensus" 16518. .16629
repeat_region	/note="L2 repeat: matches 2575. .2696 of consensus" 17143. .17347
repeat_region	/note="Aluwb repeat: matches 1. .212 of consensus" 17885. .17974
repeat_region	/note="45 copies 2 mer aa 63% conserved" 18571. .18836
repeat_region	/note="AluSx repeat: matches 43. .307 of consensus" 18978. .19325
repeat_region	/note="MERIB repeat: matches 1. .337 of consensus" 19426. .19736
misc_feature	/note="AluSx repeat: matches 1. .312 of consensus" 19634. .20175
repeat_region	/note="match: GSS: Em:AQ486219" 20127. .20304
repeat_region	/note="Aluwb repeat: matches 125. .302 of consensus" 21392. .21514
repeat_region	/note="MER63 repeat: matches 1. .745 of consensus" 21676. .21807
repeat_region	/note="MIR repeat: matches 2. .137 of consensus" 21846. .21934
repeat_region	/note="MER5A repeat: matches 3. .112 of consensus" 21927. .21978
repeat_region	/note="MER5A repeat: matches 21. .75 of consensus" 22048. .22119
repeat_region	/note="AluJb repeat: matches 2. .73 of consensus" 22120. .22166
repeat_region	/note="AluSg/x repeat: matches 179. .225 of consensus" 22174. .22408
repeat_region	/note="AluJb repeat: matches 73. .308 of consensus" 22568. .22878
repeat_region	/note="AluSx repeat: matches 15. .309 of consensus" 23908. .23946

repeat_region	/note="AluJc/FRAM repeat: matches 268. .308 of consensus" 23947. .24186
repeat_region	/note="AlusX repeat: matches 1. .250 of consensus" 24187. .24337
repeat_region	/note="FRAM repeat: matches -5. .132 of consensus" 26377. .26923
repeat_region	/note="L1MB7 repeat: matches 5618. .6169 of consensus" 27679. .27720
repeat_region	/note="21 copies 2 mer tt 78% conserved" 27728. .28038
repeat_region	/note="AlusX repeat: matches 1. .307 of consensus" 29902. .30038
repeat_region	/note="MER45C repeat: matches 107. .253 of consensus" 30165. .30254
misc_feature	/note="MER45C repeat: matches 758. .855 of consensus" complement(30228. .30612)
misc_feature	/note="match: GSS: Em:AQ131013" complement(30235. .30580)
repeat_region	/note="match: GSS: Em:AQ636615" 30488. .30843
misc_feature	/note="THE1C repeat: matches 1. .371 of consensus" 30614. .31025
repeat_region	/note="match: GSS: Em:AQ154457" 30888. .30997
repeat_region	/note="AluJc repeat: matches 3. .135 of consensus" 31331. .31351
repeat_region	/note="MER5A repeat: matches 158. .178 of consensus" 31493. .31517
repeat_region	/note="MER5A repeat: matches 1. .158 of consensus" 31841. .32029
repeat_region	/note="MER5A repeat: matches 2. .189 of consensus" 32070. .32141
repeat_region	/note="MIR repeat: matches 54. .139 of consensus" 33507. .33811
misc_feature	/note="AlusX repeat: matches 1. .310 of consensus" complement(34342. .35892)
misc_feature	/note="match: STS: Em:AF192015" 34644. .34766
repeat_region	/note="match: GSS: Em:B16649" 34733. .34939
repeat_region	/note="MIR repeat: matches 48. .262 of consensus" 35277. .35409
repeat_region	/note="FLAM_C repeat: matches 1. .133 of consensus" 36202. .36282
repeat_region	/note="MIR repeat: matches 153. .228 of consensus" 36283. .36621
repeat_region	/note="MLR1A repeat: matches 1. .365 of consensus" 36622. .36644
repeat_region	/note="MIR repeat: matches 228. .251 of consensus" 37574. .37672
repeat_region	/note="L2 repeat: matches 1814. .1925 of consensus" 37574. .37672
Query Match 2.9%; Score 42.8; DB 93; Length 195021;	
Best Local Similarity 56.3%; Pred. No. 4.2;	
Matches 80; Conservative 0; Mismatches 62; Indels 0; Gaps 0;	
QY 1336	gctttgttagaattgtgtttattgttcacaggttagagaacaccacacacagatgta 1395 
Db 2521	GTTTTTTAAATACCTATGTGTCATTTTACCAAATAGTAGAATATTTTATATCATTTGCAA 2462 
QY 1396	tctgaagagataagcttctctatgtctctaaagaatggccgatacgaataaacaagcat 1455 
Db 2461	ACTGAAAGCTGATCTCTGGCTTTCTTATATATAATACCTTTTGTGAATAAAGATAAA 2402 
QY 1456	cattaaagattaaaaaiaaa 1477 
Db 2401	CATGAATGAATAAAGACTAGA 2380 

Search completed: July 12, 2001, 02:01:22  
Job time: 6372 sec

Thu Jul 12 11:15:41 2001

us-09-545-072a-1.rge

Page 15





Db 187 RRIHQQAQVLEPPMIKERRRTNSELITKGVNGALFAIGEVLYITRPLIYVLFIRKYGVR 246  
Qy 264 SWIPWALSISVDTLGMGLANSKWGEKSKQVHFSQPEKDE-----LRRKRLIWA 313  
Db 247 SWIPWALSISVDTLGMGLANSKWGEKSKQVHFSQPEKDEVRNLDLNTTTRRRKLIWA 306  
Qy 314 LYLMDRPFPTKYTRQKLESSOKKLELPLIGLFTKIVELLEGASQRYTISGS 367  
Db 307 LYLMDRPFPTKYTRQKLESSOKKLELPLIGLFTKIVELLEGASQRYTISGS 360

RESULT 2  
C96767  
unknown protein F2P9.17 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 text\_change 31-Mar-2001  
C:Accession: C96767  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: C96767  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1273 <STO>  
A:Cross-references: GB:AE005173; NID:g7109476; PIDN:AAF36740.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: F2P9.17  
A:Map position: 1

Query Match 5.2%; Score 99; DB 2: Length 1273;  
Best Local Similarity 19.7%; Pred. No. 3.7;  
Matches 73; Conservative 43; Mismatches 105; Indels 150; Gaps 17;  
Qy 40 PEAVTALGIFTTINEHIENATPRGHVSSGNDPSLSYPLLIATLKLE----- 90  
Db 950 PEPVAA-----DETQHL-----EPR---MESQNPQSHENPIVHEIPSDVEGTELAHR 996  
Qy 91 -----TVVEVAAEHFVGDKKNYIILTEAMKAVIRLALFRNSGYKMLQ-- 135  
Db 997 EANPPTKEPQEPDVSVSVSH-----EVKKSIVIRKV-RSGATSRAECSA 1042  
Qy 136 -----GETPNEKSNQSESONRAGNSGRNIGPHGLGNQHHNPWN-- 176  
Db 1043 RTIERSQIVVRHIDRGTSASVAPQRISTDAV-----SISQNHVEEVNSC 1092  
Qy 177 --LGRAMSALES-----FGONAR-TTTSSTQWSHRIHQQAQVLEPPMIKERRT 224  
Db 1093 HDVGSRTASIGSVKFASEGIDGFKELQCTAEGSKGSTOKADNNRTVPPSLP----- 1147  
Qy 225 MSELLTEKGVNGALFAIGEVLYITRPLIYVLFIRKYGVRSWIPWALSISVDTLGMGLAN 284  
Db 1148 ----LDHSMENEAQ-----QKVA-----SLQTLSIG----- 1169  
Qy 285 SKWGEKSKQVHFSQPEKDELRRLKLIWALYLMRDPEFTKYTRQKLESSOKKLELPLIG 344  
Db 1170 ----KEKEK---KDEKKEKKKKR-----EDPVYKLEKRLKKEKKKEKMAKLVS 1214  
Qy 345 FLTEKIVELLE 355  
Db 1215 STTDPAKKIE 1225

RESULT 3

T40591  
hypothetical protein SPBC646.15c - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 04-Mar-2000  
C:Accession: T40591  
R:Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, January 1999  
A:Reference number: 221938  
A:Accession: T40591  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-376 <SEE>  
A:Cross-references: EMBL:AL03216; PIDN:CAA22819.1; GSPDB:GN00067; SPDB:SPBC646.15c  
A:Experimental source: strain 972h-; cosmid c646  
C:Genetics:  
A:Gene: SPDB:SPBC646.15c  
A:Map position: 2  
A:Introns: 49/1; 126/2; 312/2; 350/1  
C:Superfamily: Schizosaccharomyces pombe hypothetical protein SPBC646.15c

Query Match 5.0%; Score 96; DB 2: Length 376;  
Best Local Similarity 26.5%; Pred. No. 1.3;  
Matches 39; Conservative 19; Mismatches 45; Indels 44; Gaps 8;  
Qy 227 ELLTEKGVNGALFAIGEVLYITRPLIYVLF-----RK-----YGVRSWIPWALSISVDTL 277  
Db 207 EIFTKRLPLNRIFS--NFIKVCRLIYLMFMWHRKQKSSSLKVRPWPWIVAFVFEVI 264  
Qy 278 -----GMCILANSKWGEKSKQVHF-----SGPEKDELRR---RKLI 311  
Db 265 SOLIDRRCESATSSRQGFLEERT---NQSQFHVVVMAFTQGRFYDEFTKHINRSL 320  
Qy 312 WALYLMRDPFFTKYTRQKLESSOKKLE 338  
Db 321 WWSNI---PVFGKYLILLSVEERQKSLE 344

RESULT 4  
A64474  
hypothetical protein MJ1394 - Methanococcus jannaschii  
C:Species: Methanococcus jannaschii  
C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 21-Jul-2000  
C:Accession: A64474  
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak  
Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,  
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.  
Science 273, 1058-1073, 1996  
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese  
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc  
A:Reference number: A64300; MUID:96337999  
A:Accession: A64474  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-987 <BUL>  
A:Cross-references: GB:U67579; GB:L77117; NID:g1592037; PIDN:AAB99404.1; PID:g1592041  
C:Genetics:  
A:Map position: FOR1343894-1346857  
A:Start codon: GTG

Query Match 5.0%; Score 96; DB 2: Length 987;  
Best Local Similarity 20.4%; Pred. No. 4.8;  
Matches 60; Conservative 37; Mismatches 131; Indels 66; Gaps 10;  
Qy 100 FYGDKKNYIILTEAMKAVIRLALFRNSGYKMLLOGGETPNEEKDSNQSESON----- 152  
Db 577 FYGDKKNYIILTEAMKAVIRLALFRNSGYKMLLOGGETPNEEKDSNQSESON-----KDINTSNLAINWTNIT 624  
Qy 153 -RAGNSGRNIGPHGLGNQHHNPWNLEGRAMSALSSFGQNARTTTSSTPG-----WSRR 205  
Db 625 LNWNSDSATLVFNVLGNYSYERDNLAKYGFPAKILFNGTNTNTSIKGVYASGSYSIS 684



QY 206 IQHQ-----QAVIEPPIKERRRTMSSELLTEKGVNGALFAIGEVLYITRPIYVLFIRKY 260  
Db 685 TDHGTGTEINIENWTFKNDKASYFNLTNLI-----WAVNKSAYELYNWPF 733  
QY 261 GYRSWIP---WAISLSVDTLGMLLANSKWGEKSKQVHFSGPKEDELRRKKLIWA---L 314  
Db 734 NKSIMDSNVTITPNID-----IPGEVWNSKTYNFTFSGV-----PIYVANCSEF 779  
QY 315 YLMRDPFFTKYTRQKLESSQKKLELIPLIGF---LTKIVELLEGAQSRITYI 364  
Db 780 TLSKDYILLNEVSGISGVYVVEIYVVGSLYIKVHKHIVPDADGIYDIIVV 833  
RESULT 5  
C70191  
penicillin-binding protein (pbp-3) homolog - Lyme disease spirochete  
C:Species: Borrelia burgdorferi (Lyme disease spirochete)  
C:Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 08-Oct-1999  
C:Accession: C70191  
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White  
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,  
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.  
Nature 390, 580-586, 1997  
A:Authors: Smith, H.O.; Venter, J.C.  
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.  
A:Reference number: A70100; MUID:98065943  
A:Accession: C70191  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-932 <KLE>  
A:Cross-references: GB:AE001173; GB:AE000783; NID:g2688665; PIDN:AAAC67082.1; PID:g268867  
A:Experimental source: strain B31  
Query Match 5.0%; Score 95.5; DB 2; Length 932;  
Best Local Similarity 21.18; Pred. No. 4.9;  
Matches 84; Conservative 39; Mismatches 121; Indels 155; Gaps 19;  
QY 45 AFLG---IFTTINEHIENAPT-----RGHVSSGN----- 73  
Db 393 AILGIDKIDKATKEYIIRKNTLPKLIQAPEGAMIAIDTTSACIRAMVGGSGHTKDNFNR 452  
QY 74 -----DPSLSY-PLIIAILKDLTEV-----EVAEHEYGDKKNYI 109  
Db 453 ATQAKVQPSAFKALYFAAAILDKITAATMFSDSPVAFLNKNGEYVAPNGYG-KWRGN 511  
QY 110 ILTE-----AMKAVIRLALFRNSGYKMLQGGTPE-EKD-----SN 146  
Db 512 VITRALALSINIPALRILDRGFDLSAISYSKLLGITDPKEIEKTPKVPYPLALGVISV 571  
QY 147 QSEQNRA-----GNSGRNLGPHGLGNQNHNPWNLEGRAMSALSSFGQNARTTTSSTPGW 202  
Db 572 SPIQARAFAILNGSGSEIEPYGIR-----YIEDRA-----GRIITNEEASILAK 616  
QY 203 SRIHQOQAVIEP-----PMIKERRRTMSSELLTE-KGVNGAL----- 238  
Db 617 IKKEHOTQIVSPQPAYIITDMKSTIQYGTLANQRVTLNKNFKSDIAGSGTQTQNWADG 676  
QY 239 FAIGEVLYITRPIYVLFIRKGVRSWIPWAISLSVDTLGMLLANSKWGEKSKQVHFS 298  
Db 677 WAIGVSPYIT-TAFVVGFDKK-----CYSLSIGSTGTGLAGPS--WGEFMAEYHK 724  
QY 299 GPEK-----DELRRKKLIWALYL 316  
Db 725 LPKKVFKVKGAGIIPVQAETGLLPEETADEKIINELFI 763  
RESULT 6  
C83499  
probable transglycolase Pall171 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: C83499

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Miziochichi, S.D.; Warrenner, P.; Hickey, M.J.;  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L  
.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa  
A:Reference number: A82950; MUID:20437337  
A:Accession: C83499  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-398 <STO>  
A:Cross-references: GB:AE004547; GB:AE004091; NID:g9947089; PIDN:AG04560.1; GSPDB:GN  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA1171  
Query Match 4.8%; Score 93; DB 2; Length 398;  
Best Local Similarity 26.8%; Pred. No. 2.5;  
Matches 62; Conservative 26; Mismatches 85; Indels 58; Gaps 14;  
QY 155 GNSGRNL---GPHGLGNQNH-----NPNWLEGR--AMSALSSFGQNARTTTSSTP 200  
Db 198 GDGRDLVSGVPDALGASTANYLKKAGWRTGCPWGYEVKVPADFPASLAGRGRQPLSA-- 255  
QY 201 GW-----SRRTHQOQAVIEPPIKERRRTMSSELLTEKGVNGALFAIGEVLYITRPIYVLF 256  
Db 256 -WARGVRRVVDGQPL-----PGGDEK-----AAILLPAGAOGPAF-----LVRYNYDAIV--- 299  
QY 257 IRKYGVRWSWIPWAISLSVDTL--GMGLLANSKWGEKSKQVHFSGPKEDELRRKKLIWAL 314  
Db 300 --SYNAEYALAIALLSDRLRGSSGLVAS---WPT-----DDPGISRLERKKLQKAL 347  
QY 315 YLMR-----DPFTKTYTRQKLESSQKKLELIPLIGFLTEKIVELEGAQ 358  
Db 348 -LARGYDIGEADGLIGTSTRKAIQAQKRLGLTPADGRAGRKILEALKGAQ 397  
RESULT 7  
S25370  
MSB2 protein - yeast (Saccharomyces cerevisiae)  
N:Alternate names: protein G4017; protein YGR014W  
C:Species: Saccharomyces cerevisiae  
C:Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 21-Jul-2000  
C:Accession: S25370; S64305  
R:Bender, A.; Pringle, J.R.  
Yeast 8, 315-323, 1992  
A:Title: A Ser/Thr-rich multicopy suppressor of a cdc24 bud emergence defect.  
A:Reference number: S25370; MUID:92383951  
A:Accession: S25370  
A:Molecule type: DNA  
A:Residues: 1-1306 <BEN>  
A:Cross-references: GB:M77354; NID:g171993; PIDN:AAA34798.1; PID:g171994  
R:Rieger, M.; Mueller-Auer, S.; Brueckner, M.; Schaefer, M.  
submitted to the Protein Sequence Database, May 1996  
A:Reference number: S64071  
A:Accession: S64305  
A:Molecule type: DNA  
A:Residues: 1-1306 <RIE>  
A:Cross-references: EMBL:272799; NID:g1322977; PID:g1322978; MIPS:YGR014W  
A:Experimental source: strain S288C  
C:Genetics:  
A:Gene: SGD:MSB2  
A:Cross-references: SGD:S0003246; MIPS:YGR014W  
A:Map position: 7R  
C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosida  
C:Keywords: transmembrane protein  
F:3-19/Domain: transmembrane #status predicted <TM1>  
F:1189-1205/Domain: transmembrane #status predicted <TM2>  
Query Match 4.8%; Score 93; DB 2; Length 1306;  
Best Local Similarity 21.6%; Pred. No. 13;  
Matches 61; Conservative 49; Mismatches 83; Indels 90; Gaps 15;



Db 152 NDIDDNHVG1-----DLNGVTSIASAPAAAYFDDREAKNISLRASGKP 194  
 QY 119 IRLALFNSGYKMLQGGTPENEKDSNQESQNRAGNSGRNLGPHGLGNQHNPNWLE 178  
 Db 195 VRWIEYNATETMLNVTPLADRPKPSIPULS-----RKWNLGSGIFSQEH----- 240  
 QY 179 GRAMSALSSFGQARTTSS--TPGWSRRHQHQA---VIEPPMIKERRRTMS-----EL 228  
 Db 241 -----VGFSASTGVVASSHFVGLWSFNIEGKESDFDIKLSLPDPPTLSFSPSPV 293  
 QY 229 LTKGVNGALFAIGEVLITRPLIYVLFIRKYGVRS-----WIPWALSVDTL--GMGLL 282  
 Db 294 STEKSNNTMLIIIVAAASATVALMILIFSGFWFLRRDKIFFIGCARKFSYQTSNATGGF 353  
 QY 283 ANSKWGEKSKOVHFG 299  
 Db 354 DNSKLLGERNSGSFYKG 370

RESULT 11  
 T51650  
 hypothetical protein F19B10.10 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
 C:Accession: T51650  
 R:Gattung, S.  
 submitted to the EMBL Data Library, April 1997  
 A:Description: The sequence of C. elegans cosmid F19B10.  
 A:Reference number: Z18302  
 A:Accession: T51650  
 A:Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-639 <GAT>  
 A:Cross-references: EMBL:AF000261; NID:g1947114; PID:g1947124; PIDN:AA52930.1; GSPDB:CN  
 A:Experimental source: strain Bristol N2; clone F19B10  
 C:Genetics:  
 A:Gene: CESP:F19B10.10  
 A:Map position: 2  
 A:Introns: 29/3; 67/1; 101/3; 137/3; 272/3; 339/3; 407/3; 509/3

Query Match 4.7%; Score 90; DB 2; Length 639;  
 Best Local Similarity 20.5%; Pred. No. 8;  
 Matches 53; Conservative 47; Mismatches 97; Indels 62; Gaps 12;  
 QY 1 MEAYKQWNRREYVQSGFANGTLWLLPEKFSASEIG---PEAVTAL-----GIFT 51  
 Db 371 LEHLRDLFFGSRVQSPTRSMFNGYVISMNEIFNSAQGIYFLEESIEVEKISSGLFG 430  
 QY 52 TINPHIENAPTRGH-VGSSGNDPSLSYPLLIALKDLTVVVEAAEHFYGDKKNYI- 109  
 Db 431 TPESKTIRDVP-PKGFTVGDVHSE--LKYLGLLEFFFNISYAEIAYKVVAAEKMKILR 487  
 QY 110 --ILTEAMKAVIRALFRNGYKML-----LOGGETPNEEKDSNQESQNRAG 155  
 Db 488 TCDMFAVEICMLAVF--NGYPKLAQFLHSQHACHRIQGLQCKKCELEK----- 536  
 QY 156 NSGRNLGPHGLGNQHNPNWL-----EGRAMSALSSFGQARTTSSTPGWSRRHQH 209  
 Db 537 -----HPW-----WNSWLYPYMEVEDEKVDSTSSDDEEKRTSTSELAGNRKKEQK 583  
 QY 210 QAVIEPP-----MIKERRRT 224  
 Db 584 KLVKDPPEPSTMEKTRKS 602

RESULT 12  
 T00415  
 hypothetical protein H\_248015.1 - human (fragment)  
 C:Species: Homo sapiens (man)  
 C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 05-Nov-1999  
 C:Accession: T00415  
 R:Fulton, B.; Nhan, M.; Hawkins, J.; Beck, C.

submitted to the EMBL Data Library, September 1998  
 A:Description: The sequence of Homo sapiens PAC clone 248015.  
 A:Reference number: Z14148  
 A:Accession: T00415  
 A:Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1849 <FUL>  
 A:Cross-references: EMBL:AC002483; NID:g3598729; PIDN:AAC35295.1; PID:g2341020  
 C:Genetics:  
 A:Map position: 13q12-q13  
 A:Note: H\_248015.1

Query Match 4.7%; Score 89.5; DB 2; Length 1849;  
 Best Local Similarity 23.3%; Pred. No. 42;  
 Matches 48; Conservative 32; Mismatches 87; Indels 39; Gaps 10;  
 QY 76 SLSYPLLIAILKD-LETVVEAAEHFYGDKKNYIILTEAMKAVIRLA-----LFRN 126  
 Db 1280 ALKQPLSAHALSDLLSRLVEVIGEH--GDEIQGVV-----MEALLTLEAAVDNLSCLKN 1332  
 QY 127 SGYKMLQGGTPENEKDSNQESQNRAG---NSGRNLGPHGLGNQHNPNWLEGRAM 182  
 Db 1333 SOLLTVLSRSSPDLSSSKLTASRKSTGQLNMNPGTSGNTATAERSHQ-----RSF 1386  
 QY 183 SALSSFGQARTTSSTPGWSRRHQHQAIVTEPPMIKERRRTMS--ELLTEKG-VNGALF 239  
 Db 1387 SVPKFKGVDR---SSDPPRSATLDRIQACTQOGLSKTSRSSSSSLKSLDTPSHINPTN 1443  
 QY 240 AIGEVLYITRPLI-----YVLFIR 258  
 Db 1444 LLATIFWTVTALMESDFEYFLMALR 1469

RESULT 13  
 T51650  
 probable transcription factor MYB36 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 18-Aug-2000 #sequence\_revision 18-Aug-2000 #text\_change 06-Oct-2000  
 C:Accession: T51650  
 R:Kranz, H.D.; Denekamp, M.; Greco, R.; Jin, H.; Kranz, H.D.; Denekamp, M.; Greco, R.  
 ; Paz-Ares, J.; Weissshaar, B.  
 Plant J. 16, 263-276, 1998  
 A:Title: Towards functional characterisation of the members of the R2R3-MYB gene from  
 A:Reference number: Z14349; MUID:9839469  
 A:Accession: T51650  
 A:Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-333 <KRA>  
 A:Cross-references: EMBL:AF062878; PIDN:AAC83600.1  
 C:Genetics:  
 A:Gene: MYB36  
 A:Map position: 5  
 C:Superfamily: myb-related protein MIXTA; myb DNA-binding repeat homology

Query Match 4.6%; Score 89; DB 2; Length 333;  
 Best Local Similarity 17.3%; Pred. No. 4.4;  
 Matches 59; Conservative 35; Mismatches 113; Indels 134; Gaps 12;  
 QY 12 REYVQSGFSGFANGTTLW-LPEKFSASEIGPEAVTAFGLCIFTINEHIENAPTR-CHVG 69  
 Db 26 KDYIDKYGTGGN---WIALPOKIGLKRKGSCLRWL-----NYLRPNIKHGG 70  
 QY 70 SSGNDPSLSYPLLIALKDLTVVVEAAEHFYGDKKNYIILTEAMKAVIRALFRNSGY 129  
 Db 71 FSEEDRIILSLYLSI-----GSRWSIIAAQLPGRTDNDIKNYWNTKL 113  
 QY 130 KMLQGGTPENEKDSNQESQNRAGNSGRNLGPHGLGN-----QNHNNPNWLEGRAMS 183  
 Db 114 KKKLLGRQKQNRQDSITDSTENNLSNNNNKSPQNLNSALEKILQLHMQNLQ-----S 169  
 QY 184 ALSFGQN-----ARTTSSTPGWSRR-----IQHQAVIEPPMIKERRR 223







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2001, 11:35:01 ; Search time 24.54 Seconds  
(without alignments)  
1978.645 Million cell updates/sec

Title: US-09-545-072A-2  
Perfect score: 1922  
Sequence: 1 MEAYKQWNRREYVQSGFS.....EKIVELLEGAQSRVYIIGS 367

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_16.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phase.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_unclassified.\*  
13: sp\_vertebrate.\*  
14: sp\_virus.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					
Result No.	Score	Query %	Match Length	ID	Description
1	1922	100.0	367	10 Q9XEG0	Q9xeg0 arabidopsis
2	1773	92.2	360	10 Q9SLI9	Q9sl19 arabidopsis
3	249.5	13.0	336	4 Q9V5Y5	Q9v5y5 homo sapien
4	227.5	11.8	341	5 Q9VPB9	Q9vpb9 drosophila
5	97	5.0	855	10 Q9FYN9	Q9fyn9 oryza sativ
6	96.5	5.0	1957	4 O15064	O15064 homo sapien
7	96	5.0	376	3 Q94516	Q94516 schizosacch
8	95.5	5.0	932	2 O51674	O51674 borrelia bu
9	94	4.9	1005	4 O75336	O75336 homo sapien
10	93	4.8	398	2 O914G6	O914g6 pseudomonas
11	92	4.8	1196	3 Q9HDP9	Q9hdp9 cephalospor
12	91	4.7	296	2 Q9KT82	Q9kt82 vibrio chol
13	90.5	4.7	330	10 Q9M8K2	Q9m8k2 arabidopsis
14	90.5	4.7	1277	13 Q98902	Q98902 fugu rubrip
15	90	4.7	627	10 Q9ZW09	Q9zw09 arabidopsis
16	90	4.7	639	5 O02075	O02075 caenorhabdi
17	89.5	4.7	1257	2 P96746	P96746 corynebacte
18	89.5	4.7	1295	2 O34110	O34110 helicobacte
19	89.5	4.7	1849	4 O14572	O14572 homo sapien

20	89.5	4.7	3012	4 Q9Y3N6	Q9y3n6 homo sapien
21	89	4.6	333	10 Q9ZTE6	Q9zte6 arabidopsis
22	89	4.6	333	10 Q9FKL2	Q9fkl2 arabidopsis
23	89	4.6	3165	14 Q04350	Q04350 cryphonectr
24	88.5	4.6	886	11 Q9ES60	Q9es60 mus musculu
25	88.5	4.6	890	11 Q9ES59	Q9es59 mus musculu
26	88.5	4.6	1072	11 Q9R243	Q9r243 mus musculu
27	88.5	4.6	1172	11 Q9R244	Q9r244 mus musculu
28	88	4.6	390	2 P70928	P70928 bacillus li
29	88	4.6	425	10 Q9ST79	Q9st79 oryza sativ
30	87.5	4.6	320	2 Q9FAF7	Q9faf7 agrobacteri
31	86.5	4.5	285	10 Q9FT68	Q9ft68 arabidopsis
32	86.5	4.5	313	10 Q9SVM6	Q9svm6 arabidopsis
33	86.5	4.5	879	2 O31504	O31504 bacillus su
34	86.5	4.5	1025	10 Q40640	Q40640 oryza sativ
35	86	4.5	304	2 Q9REA3	Q9rea3 enterococcu
36	86	4.5	1872	5 Q17083	Q17083 athalia ros
37	85.5	4.4	649	4 Q9P215	Q9p215 homo sapien
38	85.5	4.4	784	2 Q9HXB2	Q9hxb2 pseudomonas
39	85.5	4.4	827	4 Q9ULJ0	Q9ulj0 homo sapien
40	85.5	4.4	881	3 Q9UVF5	Q9uvf5 yarrowia li
41	85.5	4.4	1290	2 O34111	O34111 helicobacte
42	85.5	4.4	2380	5 Q96266	Q96266 plasmodium
43	85	4.4	414	2 Q9Z7D0	Q9z7d0 chlamydia p
44	85	4.4	535	5 O18009	O18009 caenorhabdi
45	85	4.4	1238	5 O61198	O61198 caenorhabdi

ALIGNMENTS

RESULT 1  
Q9XEG0 PRELIMINARY; PRT; 367 AA.

AC Q9XEG0;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE SHRUNKEN SEED PROTEIN.  
OS SSEL.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
OC Brassicales; Brassicaceae; Arabidopsi.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. C24;  
RX MEDLINE=99212146; PubMed=10195899;  
RA Lin Y., Sun L., Nguyen L.V., Rachubinski R.A., Goodman H.M.;  
RT "The Pex16p homolog SSE1 and storage organelle formation in  
RT Arabidopsis seeds";  
RL Science 284:328-330(1999).  
DR EMBL; AF085354; AAD30661.1; -.  
DR InterPro; IPR000002; -.  
DR ProDom; PD004563; -; 1.  
SQ SEQUENCE 367 AA; 41624 MW; 35EE727140C485B8 CRC64;

Query Match 100.0%; Score 1922; DB 10; Length 367;  
Best Local Similarity 100.0%; Pred. No. 1.6e-166;  
Matches 367; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEAYKQWNRREYVQSGFSFANGLTWLLPEKFSASEIGPEAVTAFIGITTFINEHIEN 60  
Db 1 MEAYKQWNRREYVQSGFSFANGLTWLLPEKFSASEIGPEAVTAFIGITTFINEHIEN 60

QY 61 APTPRGHVSSGNDPSLSYPLLIILKDLKLETVVEVAEAEHFYGDKNWYIILTEAMKAVIR 120  
Db 61 APTPRGHVSSGNDPSLSYPLLIILKDLKLETVVEVAEAEHFYGDKNWYIILTEAMKAVIR 120

QY 121 LALFRNSGYKMLLQGGTPEPNEEKDSNQSESQNRAGNSGRNLGPHGLGNQNHHPNLEGR 180  
Db 121 LALFRNSGYKMLLQGGTPEPNEEKDSNQSESQNRAGNSGRNLGPHGLGNQNHHPNLEGR 180

Db 121 LALFRNSGYKMLQGGGETPNEEKDSNQSESNRAGNSGNLPGHGLGNHNPWNLEGR 180  
QY 181 AMSALSSFGQNARTTSSPGWSRRIOHQOAVIEPPMIKERRRTMSSELLTEKGVNGALFA 240  
Db 181 AMSALSSFGQNARTTSSPGWSRRIOHQOAVIEPPMIKERRRTMSSELLTEKGVNGALFA 240  
QY 241 IGEVLITRPLIYVLFIRKYGVRSPWPAISLSVDTLGMGLLANSKWMGEKSKQVHFSGP 300  
Db 241 IGEVLITRPLIYVLFIRKYGVRSPWPAISLSVDTLGMGLLANSKWMGEKSKQVHFSGP 300  
QY 301 EXDELRRRLKLIWALYMRDPFTTKYTRQKLESSQKLELIPLIGLTKIYVLELLEGAQSR 360  
Db 301 EXDELRRRLKLIWALYMRDPFTTKYTRQKLESSQKLELIPLIGLTKIYVLELLEGAQSR 360  
QY 361 YTYISGS 367  
Db 361 YTYISGS 367  
RESULT 2  
Q9SLL9 PRELIMINARY; PRT; 360 AA.  
AC Q9SLL9  
DT 01-MAY-2000 (TReMBLrel. 13, Created)  
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)  
DE AT2645690 PROTEIN.  
GN AT2645690.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
OC Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RX MEDLINE=20083487; PubMed=10617197;  
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,  
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,  
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,  
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,  
RA Copenhagen G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,  
RA Salzberg S.L., Fraser C.M., Venter J.C.;  
RT "Sequence and analysis of chromosome II of Arabidopsis thaliana";  
RL Nature 402:761-768(1999).  
DR EMBL; AC004665; AAC28524.2; -;  
SQ SEQUENCE 360 AA; 40603 MW; 79EE1446FFDADE41 CRC64;

Query Match 92.2%; Score 1773; DB 10; Length 360;  
Best Local Similarity 96.9%; Pred. No. 5.7e-153;  
Matches 343; Conservative 1; Mismatches 0; Indels 10; Gaps 1;  
QY 24 GLTWLLPEKFSASEIGPEAVTAFLGIFTTNEHIENAPTGRGHVSSGNDPSLSYPLLI 83  
Db 7 GLTWLLPEKFSASEIGPEAVTAFLGIFTTNEHIENAPTGRGHVSSGNDPSLSYPLLI 66  
QY 84 AILKDETVEVAEAEHFYGDKKWNIITTEAMKAVIRLALFRNSGYKMLQGGGETPNEEK 143  
Db 67 AILKDETVEVAEAEHFYGDKKWNIITTEAMKAVIRLALFRNSGYKMLQGGGETPNEEK 126  
QY 144 DSNQSESNRAGNSGNLPGHGLGNHNPWNLEGRAMSALSSFGQNARTTSSPGWS 203  
Db 127 DSNQSESNRAGNSGNLPGHGLGNHNPWNLEGRAMSALSSFGQNARTTSSPGWS 186  
QY 204 RRIHQOQAVIEPPMIKERRRTMSSELLTEKGVNGALFAIGEVLITRPLIYVLFIRKYGVR 263  
Db 187 RRIHQOQAVIEPPMIKERRRTMSSELLTEKGVNGALFAIGEVLITRPLIYVLFIRKYGVR 246  
QY 264 SWIPWALSISVDTLGMGLLANSKWMGEKSKQVHFSGPEKDE-----LRRKKLIWA 313  
Db 247 SWIPWALSISVDTLGMGLLANSKWMGEKSKQVHFSGPEKDEVRNLLDNTLRRKKLIWA 306

QY 314 LYLMRDFFTKYTRQKLESSQKLELIPLIGLTKIYVLELLEGAQSRITYISGS 367  
Db 307 LYLMRDFFTKYTRQKLESSQKLELIPLIGLTKIYVLELLEGAQSRITYISGS 360  
RESULT 3  
Q9Y5Y5 PRELIMINARY; PRT; 336 AA.  
AC Q9Y5Y5  
DT 01-NOV-1999 (TReMBLrel. 12, Created)  
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)  
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)  
DE PEROXISOMAL BIOGENESIS FACTOR 16.  
GN PEX16.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99122987; PubMed=9922452;  
RA South S.T., Gould S.J.;  
RT "Peroxisome synthesis in the absence of preexisting peroxisomes.";  
RL J. Cell Biol. 144:255-266(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Honsho M., Tamura S., Shimozawa N., Suzuki Y., Kondo N., Fujiki Y.;  
RT "Mutation in PEX16 is causal in the Peroxisome-deficient Zellweger  
RT Syndrome of Complementatation Group D.";  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF118240; AAD22466.1; -;  
DR EMBL; AB016531; BAA88826.1; -;  
SQ SEQUENCE 336 AA; 38643 MW; 0C0688C52E7AE47C CRC64;

Query Match 13.0%; Score 249.5; DB 4; Length 336;  
Best Local Similarity 25.4%; Pred. No. 1.8e-14;  
Matches 94; Conservative 68; Mismatches 157; Indels 51; Gaps 12;  
QY 4 YKQWVRNREYVQSGFSGFANGLTWLLPEKFSASEIGPEAVTAFLGIFTTNEHIENAPT 63  
Db 11 YQETVTRPATQAQLETAVRGFSYLLAGRADSHSELSELVYSASNLVLLNDGILRKELR 70  
QY 64 PRGHVSSGNDPSLSYPLLI---TAILKDETVEVAEAEHFYGD-KKWNITLTEAMKAVI 119  
Db 71 KKLPPV-----SLSQKLLTWSLVECFEVEFMEMGAAYKVGWVRLVIALQLAKAVL 123  
QY 120 RLALFRNSGYKMLQGGGETPNEEKDSNQSESNRAGNSGNLPGHGLGNHNPWNLEGR 179  
Db 124 RMLLL--LWFKAGLIQ--TSPPVPLDRETQAQPPDGD-----HSPGNHE- 163  
QY 180 RAMSALSSFGQNARTTSSPGWSRRIOHQOAVIEPPMIKERRRTM--SELLTEKVNGA 237  
Db 164 --QSYGKRNRNVVRLQNTP--SLHSRWGA----PQREGQQQOHHSELSATPTPLGL 215  
QY 238 LFAIGEVLITRPLIYVLFIRKYGVRSPWPAISLSVDTLGMGLLANSKWMGEKSKQVHF 237  
Db 216 QETIAEFLYIARPLHLGLGLGQSRKSWKFWLLAGVVDVTSLSLSDRKGLTR- 269  
QY 298 SGPKDELRRRLKLIWALYMRDPFTTKYTRQK-LESSQKLELIPLIGLTKIYVLELLE 356  
Db 270 ---ERRELRRTILLYYLLRSFPYDRFSEARILFLQLLADHPGVGLVTRPLMDYLP 326  
QY 357 AOSRYTYISG 366  
Db 327 WQKIYFYSWG 336

RESULT 4  
Q9VPB9 PRELIMINARY; PRT; 341 AA.  
ID Q9VPB9  
AC Q9VPB9;







[illegible]

RESULT	11	
Q9HDP9		
ID	Q9HDP9	PRELIMINARY; PRT; 1196 AA.
AC	Q9HDP9;	
DT	01-MAR-2001	(Tremblrel. 16, Created)
DT	01-MAR-2001	(Tremblrel. 16, Last sequence update)
DT	01-MAR-2001	(Tremblrel. 16, Last annotation update)
DE	AMINOADIPATE REDUCTASE ENZYME (FRAGMENT).	
GN	LYS2.	
OS	Cephalosporium acremonium (Acremonium chrysogenum).	
OC	Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Acremonium.	
OX	NCBI_TaxID=5044;	
	RN	[1]

Db	14	GGHCL-----GRMLAALKVFGSNATGIVTTDDGGSTGRIRHCQGGI---AWG 58
QY	220	ERRTWSSELLTEKGVNGALF-----AGEV-----LYIT-----RPLYLVLFIKY 260
Db	59	DTRNCINQLITSPISWMVFYFRKGAGELGNHGLNLMITALDNLNSVRPLDAINLRN- 11-
Qy	261	GVRSWIPWAISSVDTLGM-----GLLANSKW-WGEKS 292
Db	118	-----MLKVDVNILPSEHPSDLAALANDGKKWTGETS 150
RESULT	13	
Q9M8K2	ID	Q9M8K2 PRELIMINARY; PRT; 330 AA.
AC	OC	AM8K2;
DT	01-OCT-2000	(TREMBLrel. 15, Created)
DT	01-OCT-2000	(TREMBLrel. 15, Last sequence update)
DE	01-OCT-2000	(TREMBLrel. 15, Last annotation update)
F28LI.10	PROTEIN.	
OS	Arabidopsis thaliana	(Mouse-ear cress).
OC	Eukaryota;	Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC	Magnoliophyta;	eudicotyledons; core eudicots; rosidae; eurosids II;
OC	Brassicales;	Brassicaceae; Arabidopsiis.
OX	NCBL_taxonomy=3702;	
RN	[1]	
RC	SEQUENCE FROM N.A.	
RP	STRAIN=CV. COLUMBIA;	
RA	Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B., Wu D.,	
RA	Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,	
RA	Bowman C.L., White O., Nierman W.C., Fraser C.M.;	
RT	"Arabidopsis thaliana chromosome III BAC F28Li genomic sequence.";	
RL	Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.	
DR	EMBL: AC018907; AAF30309.1;	
SQ	SEQUENCE 330 AA; 37423 MW; 0907AFC06AA738CF CRC64;	
Query Match	4.78; Score 90.5; DB 10; Length 330;	
Best Local Similarity	18.3%; Pred. No. 5;	
Matches	69; Conservative 64; Mismatches 127; Indels 117; Gaps	
QY	7	WVMNRREVQSFGSFANGLTWLLPEKFSASEIGPEAVTAFLGIFTTNEHIENAPTGRG 66
Db	2	WATKSRRCAVHTSSMASG-----DVLPREFTVFLSHCSS-----ESWVTPRS 43
QY	67	HVGSGNDPSLSYPLL-----TAIL-----KDLETVVEVAEHYPGDKNWYIILTEAM 115
Db	44	Y-----YNLLPRLPFKTALLTGTTGGREWKVAMTSKOEYFFEQGMEFVADNQL 92
QY	116	K-AVIRLALFRNSGYKMLOGGETPNE-----EKDSNOSESONRAGNSCRNLGPGL 166
Db	93	KGEFLTVPFHGHKSKEYSVSYIGRDCKETRAVIQVEISDDTDENNLSHSPSNVSDSL 152
QY	167	GNQHNNPNWNLEGAMSALSFFG-----QNARTTTSSTPGWSRRHQHOQAEVPPMIKER 221
Db	153	SNDSSHSTSNVSLRSLNSDLHGDAETESDESSEYPENLPSASISVESYE-VVNTPTTSQR 211
QY	222	---RRTMSSELLTEKGVNGALFAIGEVLVITRP-----LIYVLPIRYKV 262
Db	212	SYKRKTIE-----NPHLYDDPNNVCFETCLKRKFEVLHVQAOLVKDYGL 256
QY	263	RSWIPWAISSLVDTL-GWGLLA--NSKW-----WEEKSKOVHFSGGPKDEL----- 305
Db	257	-----IFSDNVVDYIDYGKLTKATTKWADORVCINKWOKICERNOFT--ENDSILCEIL 308
QY	306	-RRRKLIA--LYLMRD 319
Db	309	RNEDKVVYAIIKHIFRD 325

ID Q9802 PRELIMINARY; PRT; 1277 AA.  
 AC Q9802;  
 DT 01-FEB-1997 (TREMBLrel. 02, Created)  
 DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE NEURAL CELL ADHESION MOLECULE L1 PRECURSOR (N-CAM L1).  
 GN LI-CAM.  
 OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthopterygii; Acanthopterygii; Percomorpha; Tetraodontiformes;  
 OC Tetraodontidae; Takifugu.  
 OX NCBI\_TaxID=31033;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RC TISSUE=BRAIN, AND MUSCLE;  
 RX MEDLINE=98147998; PubMed=9479034;  
 RA Coutelle O., Nyakatura G., Taudien S., Elgar G., Brenner S.,  
 RA Platzer M., Drescher B., Joutet M., Kenwick S., Rosenthal A.;  
 RT "The neural cell adhesion molecule L1: genomic organisation and  
 RT differential splicing is conserved between man and the pufferfish  
 RT Fugu.";  
 RL Gene 208:7-15(1998).  
 CC - FUNCTION: CELL ADHESION MOLECULE WITH A AN IMPORTANT ROLE IN THE  
 CC DEVELOPMENT OF THE NERVOUS SYSTEM. INVOLVED IN NEURON-NEURON  
 CC ADHESION, NEURITE FASCICULATION, OUTGROWTH OF NEURITES, ETC. BINDS  
 CC TO AXONIN ON NEURONS.  
 CC - SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC - ALTERNATIVE PRODUCTS: TWO FORMS; BRAIN ISOFORM (SHOWN HERE) AND  
 CC MUSCLE ISOFORM; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC - SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS 6  
 CC C2-LIKE DOMAINS. STRONG, TO DROSOPHILA NEUROGLIAN.  
 CC - SIMILARITY: CONTAINS 5 FIBRONECTIN TYPE III-LIKE DOMAINS.  
 DR EMBL; Z71926; CAA96469.1; -;  
 DR EMBL; AF026198; AAC15580.1; -;  
 DR HSSP; P20241; ICFB.  
 DR InterPro; IPR001777; -;  
 DR InterPro; IPR003006; -;  
 DR Pfam; PF00041; fn3; 5;  
 DR Pfam; PF00047; ig; 6;  
 DR PRINTS; PR00014; ENTPEIII.  
 DR SMART; SM00060; FN3; 1.  
 DR Cell adhesion; Glycoprotein; Transmembrane; Repeat; Brain;  
 KW Immunoglobulin domain; Signal; Alternative splicing.  
 FT SIGNAL 1 29  
 FT CHAIN 30 1277  
 FT DOMAIN 30 1135  
 FT TRANSMEM 1136 1156  
 FT DOMAIN 1157 1277  
 FT DOMAIN 52 146  
 FT DOMAIN 147 252  
 FT DOMAIN 253 349  
 FT DOMAIN 350 441  
 FT DOMAIN 442 533  
 FT DOMAIN 534 631  
 FT DOMAIN 632 728  
 FT DOMAIN 729 831  
 FT DOMAIN 832 932  
 FT DOMAIN 933 1032  
 FT DOMAIN 1033 1127  
 FT DISULFID 173 224  
 FT VARSPPLIC 42 47  
 FT YRSPFLIC 1190 1193  
 FT CARBOHYD 317 317  
 FT CARBOHYD 503 503  
 FT CARBOHYD 520 520  
 FT CARBOHYD 531 531  
 FT CARBOHYD 794 794  
 FT CARBOHYD 839 839  
 FT CARBOHYD 1035 1035  
 FT CARBOHYD 1046 1046  
 FT CARBOHYD 1068 1068  
 FT CARBOHYD 1083 1083

FT CARBOHYD 1108 1108 N-LINKED (GLCNAC...) (POTENTIAL).  
 SQ SEQUENCE 1277 AA; 141954 MW; 284BB49BA4A42C27 CRC64;  
 Query Match 4.7%; Score 90.5; DB 13; Length 1277;  
 Best Local Similarity 21.5%; Pred. No. 36;  
 Matches 68; Conservative 38; Mismatches 116; Indels 95; Gaps 15;  
 QY 31 EKFSASEIGPEATAFIGTTFINEHIIENAPTPRGHVSGSGNDPSLSYPLLIALLKDL 90  
 Db 793 QNFAAEKQVAVNK-KGL-----GPEDDPIIGYGEDVPLEAPNLGVLENS 840  
 QY 91 TVVEVAAEHFYGDKKWNYIILTEAMKAVIRLALFRNSGYKMLQGGTPEEKDNOSES 150  
 Db 841 TTRVT-----WSAV-----DKETVRGHLL---GKYILTWG---HHRNSRAQEP 879  
 QY 151 QNRA-GNSGRNLGPHGLGNQNHHPNLEGRANSALSFQGN--ARTTSTSGWSRRIQ 207  
 Db 880 ENIVMVGOTGANEKKSTINLRPYCHYDL---AISAFKSGEGPLSEKTSFWTP----- 929  
 QY 208 HQQAVIEPPMIKERRRTMSELLT-----EKGVNGALFAIG-----EVLYITR 249  
 Db 930 --EGVPGPPMSKQMTSPSESEITLHWTPPKPKNGLLGYSLQYKRMOSDDNPLQVVDIAS 987  
 QY 250 PLIYVLFIRKYGVRSWIPWAI-----SLSVDTLGMGLL-----ANS 285  
 Db 988 FEITHLTLLKGLDRSHYQILMARTAAAGKGLSIEILGATTLLEGLPANISLSAEERSVNL 1047  
 QY 286 KWWGEKSK-----QVHF 297  
 Db 1048 SWEARKRHRTVGFQIH 1064  
 RESULT 15  
 Q9ZW09 PRELIMINARY; PRT; 627 AA.  
 ID Q9ZW09  
 AC Q9ZW09;  
 DT 01-MAY-1999 (TREMBLrel. 10, Created)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE PUTATIVE RECEPTOR-LIKE PROTEIN KINASE.  
 GN F16P2.40  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsi.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Rounsley S.D., Lin X., Kaul S., Shea T.P., Fujii C.Y., Mason T.M.,  
 RA Shen M., Ronning C.M., Fraser C.M., Somerville C.N., Venter J.C.;  
 RT "Arabidopsis thaliana chromosome II BAC F16P2 genomic sequence.";  
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC004561; AAC95213.1; -;  
 DR HSSP; P05045; 1BQJ.  
 DR InterPro; IPR000719; -;  
 DR InterPro; IPR001220; -;  
 DR Pfam; PF00069; pkinase; 1.  
 DR Pfam; PF00139; lectin\_legB; 1.  
 DR PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.  
 DR SMART; SM00221; STYK; 1.  
 KW ATP-binding; Kinase; Receptor; Transferase.  
 SQ SEQUENCE 627 AA; 68788 MW; 91D9F8100BEIAB1 CRC64;  
 Query Match 4.7%; Score 90; DB 10; Length 627;  
 Best Local Similarity 20.8%; Pred. No. 14;  
 Matches 66; Conservative 46; Mismatches 131; Indels 74; Gaps 11;  
 QY 16 QSFQGFANGTLWLPKFSASEIGPEAVTAFIGTTFINE-----HIEN 60  
 Db 95 QGTGAPGHGLAFVISPDMDFSGAFP---SNYLGLENTSNGNSLNRLAIEFDTVQAVEL 151

QY 61 APTPRGHVSSGNDPSLSYPLLIALLKLETVVEVAA--EHFYGDKKKXNYIILTEAMKAV 118  
Db 152 NDIDONHVGI-----DLNGVISIASAPAYFDDREAKNISLRLASGKP 194  
QY 119 IRLALFRNSGYKMLLOGGETPNEEKDSNQSESONRAGNSGRNLGPHGLGNQNNHHNPWNLE 178  
Db 195 VRVWIEYNATETMLNVTIAPLDRPKPSIPLLS-----RKMNLSGIFSQEH----- 240  
QY 179 GRAMSALSSFGONARTTTSS--TPGWSRRIOHQA---VIEPPMIKERRRTMS-----EL 228  
Db 241 -----VGFSASTGTVASSHFVLGWSFNIEGKESDFITKLPSPDPPTLSPSPSPV 293  
QY 229 LTERGVNGALFAIGEVLVITRPLIYVLFIRKYGVR-----WIPWAIISLSDTL--GMGLL 282  
Db 294 STEKKSNTMLIIIIAASATVALMILIFSGFWFLRRDKIFFIGGARKFSYQTISNATGGF 353  
QY 283 ANSKWGEKSKQVHFG 299  
Db 354 DNSKLLGERNSGSFYKG 370

Search completed: July 11, 2001, 11:37:04  
Job time: 123 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2001, 11:35:01 ; Search time 12.29 Seconds  
(without alignments)  
601.553 Million cell updates/sec

Title: US-09-545-072A-2

Perfect score: 1922  
Sequence: 1 MEAYKQWNRREYVQSFSGS.....EKIVELLEGAQSRYYTISGS 367

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 20144635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*

- 1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*
- 2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*
- 3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*
- 4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*
- 5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*
- 6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	89.5	4.7	1257	2	US-08-750-152A-2
2	89	4.6	3165	2	US-08-459-146-3
3	89	4.6	3165	2	US-08-459-065-3
4	88	4.6	1489	6	5183745-2
5	87.5	4.6	302	2	US-08-893-853-3
6	87.5	4.6	302	4	US-09-113-921-3
7	86.5	4.5	1012	2	US-08-475-891A-4
8	86.5	4.5	1025	2	US-08-567-375-4
9	86.5	4.5	1025	2	US-08-587-680A-4
10	85.5	4.4	2595	4	US-09-036-987A-2
11	81.5	4.2	1250	1	US-08-441-139-9
12	80.5	4.2	524	1	US-08-447-500-24
13	80.5	4.2	524	1	US-08-454-097-24
14	80.5	4.2	524	1	US-08-453-866-24
15	80.5	4.2	524	3	US-08-185-359-24
16	80.5	4.2	747	2	US-08-895-522-1
17	80.5	4.2	747	3	US-09-195-391-1
18	80	4.2	730	2	US-08-696-944-2
19	79.5	4.1	821	1	US-07-928-464-2
20	79.5	4.1	821	1	US-08-003-311B-2
21	79.5	4.1	821	1	US-08-261-432-2
22	79.5	4.1	821	5	PCT-US93-07347-2
23	79	4.1	1794	6	5183745-6
24	78.5	4.1	513	1	US-08-464-340A-2
25	78.5	4.1	513	5	PCT-US94-08449A-2
26	77	4.0	525	6	5183745-4
27	77	4.0	1119	4	US-09-396-651B-2

28	76.5	4.0	392	4	US-09-416-050A-2	Sequence 2, Appli
29	76.5	4.0	392	4	US-09-664-800-2	Sequence 2, Appli
30	76.5	4.0	392	4	US-09-665-309-2	Sequence 2, Appli
31	76.5	4.0	1989	4	US-08-836-325-12	Sequence 12, Appli
32	76.5	4.0	3079	5	PCT-US94-00198-4	Sequence 4, Appli
33	75	3.9	3025	6	5223423-3	Patent No. 5223423
34	74.5	3.9	307	2	US-08-540-804-6	Sequence 6, Appli
35	74.5	3.9	307	2	US-08-218-265-6	Sequence 6, Appli
36	74.5	3.9	307	3	US-08-521-872-6	Sequence 6, Appli
37	74.5	3.9	307	4	US-08-590-399-6	Sequence 6, Appli
38	74.5	3.9	1052	4	US-09-255-502-7	Sequence 7, Appli
39	74.5	3.9	1261	4	US-09-208-742-4	Sequence 4, Appli
40	74	3.9	494	1	US-08-447-500-2	Sequence 4, Appli
41	74	3.9	494	1	US-08-454-097-2	Sequence 2, Appli
42	74	3.9	494	1	US-08-447-408-2	Sequence 2, Appli
43	74	3.9	494	1	US-08-453-866-2	Sequence 2, Appli
44	74	3.9	494	3	US-08-185-359-2	Sequence 2, Appli
45	73.5	3.8	230	4	US-09-091-219-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1  
US-08-750-152A-2  
; Sequence 2, Application US/08750152A  
; Patent No. 5977331  
; GENERAL INFORMATION:  
; APPLICANT: ASAKURA, YOKO  
; APPLICANT: KIMURA, EIICHIRO  
; APPLICANT: ABE, CHIZU  
; APPLICANT: KAWAHARA, YOSHIO  
; APPLICANT: NAKAMATSU, TSUYOSHI  
; TITLE OF INVENTION: ALPHA-KETOGLUTARATE DEHYDROGENASE GENE  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.  
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR  
; CITY: ARLINGTON  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22152  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: FastSEQ Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/750,152A  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: NORMAN F. OBLON  
; REGISTRATION NUMBER: 24,618  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-413-3000  
; TELEFAX: 703-413-2220  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1257 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-750-152A-2

Query Match 4.7%; Score 89.5; DB 2; Length 1257;  
Best Local Similarity 20.2%; Pred. No. 1.3;  
Matches 62; Conservative 37; Mismatches 123; Indels 85; Gaps 13;





```
; LENGTH: 3165 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Endothia parasitica (Cryphonectria)
; STRAIN: Ep713
; US-08-459-065-3

Query Match 4.6%; Score 89; DB 2; Length 3165;
Best Local Similarity 20.3%; Pred. No. 6.2;
Matches 82; Conservative 49; Mismatches 149; Indels 124; Gaps 18;

QY 4 YKOW--VVRNREYVQSGFANGLTWLLPEKFSASEIGPEAVTAF-----LGIFTT 52
Db 1757 FEQPEVFADRDIM-----LPKGVELYIKEKISA--GTFPFISSFYKSRKALKQAGYMDV 1808
QY 53 INEHIENAPTGRGHVGGNDPSSLSY-----PLLIAILKDLTVEVAAAEHF 100
Db 1809 IRKNALECI-----STCKYPTQFYHAFKASQAVPGQPLLAPRMKDLRTVVSSELSAY 1860
QY 101 YGDKKNYIILTEAMKAVIRLALFRNSGYKMLGGTTPNEEKDSNQSQN-----RAGN 156
Db 1861 MVDQ-----IFQIEANKRITWETYGAGSGMPLSQSMARINWDELHDLRKREGGQFIADATA 1916
QY 157 SGRNLGP---HG-----LCGNHNNHNNLEGRAMSALSSFG---ONARTTTSSTPGWS 203
Db 1917 YDSNCKPALFHGAGKLVLFQNHPSG---KGROFAOVVOCKFEAMQNAVVMGITEFSYT 1973
QY 204 RRIHQOQAVIEPPMKERR-----RTMSELLTEKGVNGALFAIGEVLVYITRPLIYVLFIRK 259
Db 1974 ALTPHVDVAVRHELESKYPAHFAFESSELLAHNNVN-----VTE 2012
QY 260 YGVRSW-----IPWALSVDLTGMLLANSKMWGEKSKOVHFGSGPEKDLRR 307
Db 2013 WKRLSWEERKACARDMQAVFGKVFELTNDPALR---LOGSSWQSGFTTE-----PKRDE--- 2062
QY 308 RKLIALYLMRDPFTKYTRQKLESSQKLELIPLIGLFTKIV 351
Db 2063 -----FRKYQYFYDSKAMREDIKRIVFANREVI 2092

RESULT 4
; Patent No. 5183745
; APPLICANT: DANCHIN, ANTOINE; CLASER, PHILLIPPE; KRIN, EVELYN;
; BARZU, OCTAVIEN; LADANT, DANIEL; ULLMAN, AGNES
; TITLE OF INVENTION: ADENYL CYCLASE DERIVATIVES AND THEIR
; BIOLOGICAL USES
; NUMBER OF SEQUENCES: 13
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/426,541
; FILING DATE: 25-OCT-1989
; SEQ ID NO: 2
; LENGTH: 1489
5183745-2

Query Match 4.6%; Score 88; DB 6; Length 1489;
Best Local Similarity 21.4%; Pred. No. 2.4;
Matches 69; Conservative 40; Mismatches 105; Indels 108; Gaps 16;

QY 25 LTWLLPEKFSASEIGPEAVTAF-----LCGIFTINEHIENAPTGRGHV----- 68
Db 241 LLWKIARAGARSAYGTEARRQFYDGMNIGVITDF--ELEVNRNALNRRRAHVAAGQDVVQH 299
QY 69 GSSGNDPSSLSYPLLIIAILKDLTVEVAAAEHFYGDKNWYIILTEAMKAVIRLALFRNSG 128
Db 300 GTEONNP---FP-----EAEKIFVVSATGESG 324

; LENGTH: 3165 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Endothia parasitica (Cryphonectria)
; STRAIN: Ep713
; US-08-459-065-3

Query Match 4.6%; Score 87.5; DB 2; Length 302;
Best Local Similarity 24.1%; Pred. No. 0.22;
Matches 45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;

QY 64 PRGHVGGSGNDPSSYPLLIIAILKDLTVEVAAAEHFYGDKNWYIILTEAMKAVIRLAL 123
Db 91 PWEHPGSSGVDPRL-----EPWNHL-----GSSGVDRHLEP 121
QY 124 FRNSGYKMLGGTTPNEE-----KDSNQSQNRAKNSG---RNLPGLHGN-QN 170
Db 122 WKHPGSGDLQRRTTPQDSGSRQRRRPQDSGSRQRRRPQDSGSRQRPPGSGSKOR 181
QY 171 HHNPWNLEG-----RAMSALSFGONARTTTSSTPGWSRRIHQOQAVIEPP 216
Db 182 RRPQNSGRRRRSPQDSGSRQRRRSPQNSGGRORRRTPQSSGRRRRRAHNS----- 236

; LENGTH: 3165 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Endothia parasitica (Cryphonectria)
; STRAIN: Ep713
; US-08-459-065-3

Query Match 4.6%; Score 89; DB 2; Length 3165;
Best Local Similarity 20.3%; Pred. No. 6.2;
Matches 82; Conservative 49; Mismatches 149; Indels 124; Gaps 18;

QY 4 YKOW--VVRNREYVQSGFANGLTWLLPEKFSASEIGPEAVTAF-----LGIFTT 52
Db 1757 FEQPEVFADRDIM-----LPKGVELYIKEKISA--GTFPFISSFYKSRKALKQAGYMDV 1808
QY 53 INEHIENAPTGRGHVGGNDPSSLSY-----PLLIAILKDLTVEVAAAEHF 100
Db 1809 IRKNALECI-----STCKYPTQFYHAFKASQAVPGQPLLAPRMKDLRTVVSSELSAY 1860
QY 101 YGDKKNYIILTEAMKAVIRLALFRNSGYKMLGGTTPNEEKDSNQSQN-----RAGN 156
Db 1861 MVDQ-----IFQIEANKRITWETYGAGSGMPLSQSMARINWDELHDLRKREGGQFIADATA 1916
QY 157 SGRNLGP---HG-----LCGNHNNHNNLEGRAMSALSSFG---ONARTTTSSTPGWS 203
Db 1917 YDSNCKPALFHGAGKLVLFQNHPSG---KGROFAOVVOCKFEAMQNAVVMGITEFSYT 1973
QY 204 RRIHQOQAVIEPPMKERR-----RTMSELLTEKGVNGALFAIGEVLVYITRPLIYVLFIRK 259
Db 1974 ALTPHVDVAVRHELESKYPAHFAFESSELLAHNNVN-----VTE 2012
QY 260 YGVRSW-----IPWALSVDLTGMLLANSKMWGEKSKOVHFGSGPEKDLRR 307
Db 2013 WKRLSWEERKACARDMQAVFGKVFELTNDPALR---LOGSSWQSGFTTE-----PKRDE--- 2062
QY 308 RKLIALYLMRDPFTKYTRQKLESSQKLELIPLIGLFTKIV 351
Db 2063 -----FRKYQYFYDSKAMREDIKRIVFANREVI 2092

RESULT 5
; Sequence 3, Application US/08893853
; Patent No. 5891994
; GENERAL INFORMATION:
; APPLICANT: Goldstein, Gideon
; TITLE OF INVENTION: Methods and Compositions for Impairing
; TITLE OF INVENTION: Multiplication of HIV-1
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr., P.O. Box 457
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/893,853
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: GGP20USA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 302 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-893-853-3
```

QY 217 MIKERR 223  
Db 237 GSRQRR 243

## RESULT 6

US-09-113-921-3

Sequence 3, Application US/09113921

Patent No. 6193981

GENERAL INFORMATION:

APPLICANT: Goldstein, Gideon

TITLE OF INVENTION: Methods and Compositions for Impairing

TITLE OF INVENTION: Multiplication of HIV-1

NUMBER OF SEQUENCES: 124

CORRESPONDENCE ADDRESS:

ADDRESSEE: Howson and Howson

STREET: Spring House Corporate Cntr., P.O. Box 457

CITY: Spring House

STATE: PA

COUNTRY: USA

ZIP: 19477

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/113,921

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/893,853

FILING DATE: 11-JUL-1997

ATTORNEY/AGENT INFORMATION:

NAME: Bak, Mary E.

REGISTRATION NUMBER: 31,215

REFERENCE/DOCKET NUMBER: GGP2AUSA

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-540-9200

TELEFAX: 215-540-5818

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 302 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-113-921-3

Query Match 4.6%; Score 87.5; DB 4; Length 302;  
Best Local Similarity 24.1%; Pred. No. 0.22;  
Matches 45; Conservative 19; Mismatches 62; Indels 61; Gaps 7;

QY 64 PRGHVSSGNDPSLSYPLLIATILKDLTVVVAEAEHYGDKKNYIILTEAMKAVIRLA 123

Db 91 PWEHPSGSGVDPLR-----EPWNHL---GSSGVDHRLPP 121

QY 124 FRNSGYKMLLQGGCTPNEE-----KDSNOSSESONRAGNSG---RNLPGLHGN-ON 170

Db 122 WKHPGSDLRQRRTPDQSGSRQRRPPQDSSGRQRRPPQSGSRQRRPPQSGSROR 181

QY 171 HHNPWNLEG-----RAMSALSSFGONARTTTSTPGWSSRRHQHQAIEPP 216

Db 182 RRPQNSGRRQRRSPQDSSGGRQRRSPQNSGGRQRRRTTPQSSGGRQRRRAHNS----- 236

QY 217 MIKERR 223

Db 237 GSRQRR 243

## RESULT 7

US-08-475-891A-4

Sequence 4, Application US/08475891A  
Patent No. 5859339  
GENERAL INFORMATION:  
APPLICANT: Ronald, Pamela C.  
APPLICANT: Wang, Guo-Liang  
APPLICANT: Song, Wen-Yuang  
TITLE OF INVENTION: Procedures and Materials for Conferring  
TITLE OF INVENTION: Disease Resistance in Plants  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/475,891A  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/373,375  
FILING DATE: 17-JAN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Bastian, Kevin L.  
REGISTRATION NUMBER: 34,774  
REFERENCE/DOCKET NUMBER: 02370-0589100S  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1012 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..1012  
OTHER INFORMATION:  
OTHER INFORMATION: /note= "Xa21 Xanthomonas spp.  
disease resistance polypeptide RPK-B  
from rice (Oryza sativa)"  
US-08-475-891A-4

Query Match 4.5%; Score 86.5; DB 2; Length 1012;  
Best Local Similarity 23.9%; Pred. No. 1.9; Indels 101; Gaps 24;  
Matches 93; Conservative 50; Mismatches 145;

QY 8 VWRNREYVQSGFSGFANGTLWLLPEKFSASEIGPEAVTAFLGIFTTINEHIE--NAFTPR 65

Db 244 IW-NLSSLRASFVRENKLGMP-----TNAFKTL--HLEVIDMGTNR 284

QY 66 GHVSSGNDPSLSYPLLIATILKDLTVVVAEAEHYGDKKNYIILTEA---MKAVIRLA 122

Db 285 FHG-----KIPASVANASHL-TVIQYGNLFSG-----IITSGFRLNLTLY 327

QY 123 LFRNSGYKMLLQGGCTPNEEKDSNOSSESONRAGNSGRNLPGLHGNQHNHPWNLEGRAM 182

Db 328 LWRN-----LFQ---TREQDDWGFISDLTNCISKLOTNLGNNLGGVLPNSFSNLS----- 375

QY 183 SALSFGONARTTTSTP---GWSRRHQHQAIEP-----PMIKERRRTMSELLT-EKG 233

Db 376 TSLFLALELNKITGSIPTKIDGNLGLQLHYLCNNFRGSLPSSLGRKLNGLLAYENN 435

QY 234 VNGAL-FAIGEVLYITRPLIYVLEIRKYGVRSWIPWAIISLSDVTGLGNLANSKWGEKS 292

Db 436 LSGSIPLAIGN---LTELNILLGTNKFSGWIPYTLNLTNLLSLGSLTN-----482

QY 293 KOVHFSGPEKDELRRKLIWALYLMRDPFTKYTRQKLESS--QKKLELPLIGF-----345

Db 483 ---NLSGPIPSSEFN---IOTLSIM-----INVSNNLEGSIPQEIHLKLNLFVEFHAESN 531

QY 346 -LTKIVELLEGAG-SRYTY-----ISGS 367

Db 532 RLSGKIPNTLGDCLLYLYLQNNLLSGS 560

RESULT 8

US-08-567-375-4

; Sequence 4, Application US/08567375

; Patent No. 5952485

; GENERAL INFORMATION:

; APPLICANT: Ronald, Pamela C.

; APPLICANT: Wang, Guo-Liang

; APPLICANT: Song, Wen-Yuang

; APPLICANT: Szabo, Veronique

; TITLE OF INVENTION: Procedures and Materials for Conferring

; TITLE OF INVENTION: Disease Resistance in Plants

; NUMBER OF SEQUENCES: 16

; CORRESPONDENCE ADDRESS:

; ADDRESSER: Townsend and Townsend and Crew LLP

; STREET: Two Embarcadero Center, Eighth Floor

; CITY: San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94111-3834

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/567,375

; FILING DATE: 04-DEC-1995

; CLASSIFICATION: 800

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/004,645

; FILING DATE: 29-SEP-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/475,891

; FILING DATE: 07-JUN-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/373,375

; FILING DATE: 17-JAN-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Bastian, Kevin L.

; REGISTRATION NUMBER: 34,774

; REFERENCE/DOCKET NUMBER: 023070-058930

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 576-0200

; TELEFAX: (415) 576-0300

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1025 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-567-375-4

Query Match 4.5%; Score 86.5; DB 2; Length 1025;

Best Local Similarity 23.9%; Pred. No. 2;

Matches 93; Conservative 50; Mismatches 145; Indels 101; Gaps 24;

QY 8 VWRNREYVQSGFANGITWLLPEKFSASEIGPEAVTAFLGIFTTNEHIE--NAPTPR 65

Db 244 IW-NLSSLRASFVRENKLGMP-----TNAFKTL--HLEVIDMGTR 284

QY 66 GHVGSSGNDPSLSYLLIAIKDLLETVEVAAEHFYGKKWNYILITEA---MKAVIRLA 122

Db 285 FHG-----KIPASVANASHL-TVIQIYGNLFSG-----IITSGFGLRLNLTLY 327

QY 123 LFRNSYKMLLQGGTTPNEKDSQESQNRAGNSGRNLGPHCLGNQHHNPWNLEGRAM 182

Db 328 LWRN-----LFO---TREQDDWGFISDLTNCCKLQTLNGLNGLGVLPNSFSNLS-----375

QY 183 SALSSEFGQARTTTSTP---GMSRRIQHQAVIEP-----PMKERRRTMSELIT-EKG 233

Db 376 TSLFLALELNKLTGSIKIPKIDIGNLIGLOHLYLCNNFRGSLPSSGLRKLNLGILLAYENN 435

QY 234 VNGAL-FAIGEVLYITRPLIYVLFIRKYGVRSWIPRAISLSDVTGLMGLLANSKWWGEKS 292

Db 436 LSGSIPLAIGN---LTELNILLGTNKFSGWIPYTLNLTNLLSLGSLTN-----482

QY 293 KOVHFSGPEKDELRRKLIWALYLMRDPFTKYTRQKLESS--QKKLELPLIGF-----345

Db 483 ---NLSGPIPSSEFN---IOTLSIM-----INVSNNLEGSIPQEIHLKLNLFVEFHAESN 531

QY 346 -LTKIVELLEGAG-SRYTY-----ISGS 367

Db 532 RLSGKIPNTLGDCLLYLYLQNNLLSGS 560

RESULT 9

US-08-587-680A-4

; Sequence 4, Application US/08587680A

; Patent No. 5977434

; GENERAL INFORMATION:

; APPLICANT: Ronald, Pamela C.

; APPLICANT: Wang, Guo-Liang

; APPLICANT: Song, Wen-Yuang

; APPLICANT: Szabo, Veronique

; TITLE OF INVENTION: Procedures and Materials for Conferring

; TITLE OF INVENTION: Disease Resistance in Plants

; NUMBER OF SEQUENCES: 27

; CORRESPONDENCE ADDRESS:

; ADDRESSER: Townsend and Townsend and Crew LLP

; STREET: Two Embarcadero Center, Eighth Floor

; CITY: San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94111-3834

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/587,680A

; FILING DATE: 17-JAN-1996

; CLASSIFICATION: 800

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/373,375

; FILING DATE: 17-JAN-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/475,891

; FILING DATE: 07-JUN-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/004,645

; FILING DATE: 29-SEP-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/567,375

; FILING DATE: 04-DEC-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Bastian, Kevin L.

; REGISTRATION NUMBER: 34,774

; REFERENCE/DOCKET NUMBER: 023070-058940US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 576-0200

; TELEFAX: (415) 576-0300

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

LENGTH: 1025 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-587-680A-4

Query Match 4.5%; Score 86.5; DB 2; Length 1025;  
Best Local Similarity 23.9%; Pred. No. 2;  
Matches 93; Conservative 50; Mismatches 145; Indels 101; Gaps 24;

QY 8 VVNRREYVQSGFANGLTWLLPEKFSASEIGPEAVTAFLGIFTTINEHIETNAPTTPR 65  
DB 244 IW-NLSSRAFSVRKNKGGMIP-----TNAFKTL--HLEVIDMGNTNR 284  
QY 66 GHVSSGNDPSLSYPLLIALLKDLTVEVAAAEHFYGDKNWYIILTEA---MKAVIRLA 122  
DB 285 FHG-----KIPASVANASHL-TVIQIYNLFSG-----IISGGRRLNLTLEY 327  
QY 123 LFRNSGYKMLQGGETPNEEKDSNQESQNRAGNSGRNLGPHGLGNQHHNPNWNLGRAM 182  
DB 328 LWRN-----LFQ-----TREDDMGFISDITNGSKLQTLNLGNNLGGVLPNSFSNLS--- 375  
QY 183 SALSSFGQARTTTSSTP---GWSRRHQHQAVIEP-----PMIKERRTMSSELLT-EKG 233  
DB 376 TSLFSLALELNKITSIPKIDIGNLIGLOHLYLCNNNFRGSLPSSLGRLKNLIGILLAYENN 435  
QY 234 VNGAL-FAIGEVLYITRPLIYVIRKYGVRSWIPWALSISVDPLGMLLANSKWKMGES 292  
DB 436 LSGSIPLAIGN---LTELINLLGTNKFSGWIPYTLNLTLLSLGLSTN----- 482  
QY 293 KOVHFGSGPEKDLRRRLINWALYMRDPFTKTKTQKLESS--QKLELPLIGF----- 345  
DB 483 ---NLSGPISELEN---IQTLSTN-----INVSNNLEGSIPQEIHLKNLVEFHAESN 531  
QY 346 -LTEKIVELLEGAQ-SRITY-----ISGS 367  
DB 532 RLSGKIPTLGDCCOLLRLYLQNNLLSGS 560

RESULT 10  
US-09-036-987A-2  
Sequence 2, Application US/09036987A  
Patent No. 6143526  
GENERAL INFORMATION:  
APPLICANT: Baltz, Richard H.  
APPLICANT: Broughton, Mary C.  
APPLICANT: Crawford, Kathryn P.  
APPLICANT: Madduri, Krishnamurthy  
APPLICANT: Merlo, Donald J.  
APPLICANT: Treadway, Patti J.  
APPLICANT: Turner, Jan R.  
APPLICANT: Waldron, Clive  
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
TITLE OF INVENTION: Production  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSSEE: Dow Agrosciences LLC Patent Department  
STREET: 9330 Zionsville Road  
CITY: Indianapolis  
STATE: Indiana  
COUNTRY: USA  
ZIP: 46268  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/036, 987A  
FILING DATE: 09-MAR-1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:

NAME: Stuart, Donald R.  
REGISTRATION NUMBER: 28,479  
REFERENCE/DOCKET NUMBER: 50,608  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (317)337-4816  
TELEFAX: (317)337-4847  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2595 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-036-987A-2

Query Match 4.4%; Score 85.5; DB 4; Length 2595;  
Best Local Similarity 19.3%; Pred. No. 11;  
Matches 78; Conservative 51; Mismatches 138; Indels 137; Gaps 17;

QY 7 VVNRREYVQSGFANGLTWLLP-EKFSASEIGPEAVTAFLGIFTTINEHIETNAPTTPR 65  
DB 1766 WVRQVREPVR---FADGVOALVEHDVATVVELGPDGALSAL-----IQECVAAS 1811  
QY 66 GHVSSGNDPSLSYPLLIALLKDLTVEVAAAEHFY---GDKKW-NYIILTEAMKAVIRL 121  
DB 1812 DRAGRLSAVPMRRN-----QDEAKVMTALAHVHVRGAVDWRSFAGTRAKQIELPT 1865  
QY 122 ALFRNSGYKMLQGGETPNEEKDSNQESQNRAGNSGRNLGPHGLGNQHHNPNWNL---- 177  
DB 1866 YAFQQRQRYWL-----NALRESS-----AGDMGRVREAKFWGAVEHEDVESLARVL 1910  
QY 178 ---EGRAMSALSSSF-----GQNARTTTS----- 197  
DB 1911 GIVDDGAADVDSLSALPVLGAGWQRTTETESINDPCRYLGNRQVAGLPPMGTVFGTWLVF 1970  
QY 198 STPGHSR-----RIHQQAVI-----EPPMIKERRRT-----MSELLT 230  
DB 1971 APHWSSEPEVVDCVTALRKARGASVYLVEADPDPTSFQDRVTRTLCGLPDLVGVLMSLCL 2030  
QY 231 EKGVNGALFAIGEVLYITRPLIYVIRKYGVRSWIPWALSISV-----DTLGMGL- 281  
DB 2031 EESVLPGFSAVSRGFALTVELVRLRAAGATARLMLTCCGGVSGVDVPRPAQALAWGLG 2090  
QY 282 -----LANSKWKG-----EKSQVHFGSGPEKDELRRR 308  
DB 2091 RVVGLHPDMWGGLLIDIPVLEDEDAQERLSIVLAGLDEDEVAIR 2134

RESULT 11  
US-08-441-139-9  
Sequence 9, Application US/08441139  
Patent No. 5773245  
GENERAL INFORMATION:  
APPLICANT: Wittrup, Dr. Karl D.  
APPLICANT: Robinson, Anne S.  
TITLE OF INVENTION: METHODS FOR INCREASING SECRETION OF  
TITLE OF INVENTION: RECOMBINANTLY EXPRESSED PROTEINS  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSSEE: SCULLY, SCOTT, MURPHY & PRESSER  
STREET: 400 Garden City Plaza  
CITY: Garden City  
STATE: NY  
COUNTRY: USA  
ZIP: 11530  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/441,139  
FILING DATE: 15-MAY-1995

```

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/008,001
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Wetherell Ph.D., John R.
; REGISTRATION NUMBER: 31,678
; REFERENCE/DOCKET NUMBER: PD-2458
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 455-5100
; TELEFAX: (619) 455-5110
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 524 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-447-500-24

Query Match 4.2%; Score 80.5; DB 1; Length 524;
Best Local Similarity 22.3%; Pred. No. 3.2;
Matches 33; Conservative 17; Mismatches 61; Indels 37; Gaps 4;

Qy 74 DPSLSYPLIIALLKDLTVVVAEAEHFY-----GDKKNYIIITLTKAVIRLALFRNS 127
Db 305 DPQDYD--LISLMDALRLDLKDDGHYDMDLNGGKGNKIKINRRAN----- 350
Qy 128 GYKMLQGETPNEEKSDNQSESNRAGNSGRNLGPHGLGNQNHNPWNLEGRAMSALSS 187
Db 351 -----LHGYNPNRVNGNTARNVNTNSKTRNTTPVATPKQAQNSYNKD----- 396
Qy 188 FQGNARTTTSPTGWSRRIHQQAIVIEP 215
Db 397 ---NSKSRISNPQSFTKQHVLLKIEP 421

RESULT 13
US-08-454-097-24
; Sequence 24, Application US/08454097
; Patent No. 5686412
; GENERAL INFORMATION:
; APPLICANT: Hoekstra, Merl F.
; TITLE OF INVENTION: Protein Kinases
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/454,097
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/185,359
; FILING DATE: 21-JAN-1994
; APPLICATION NUMBER: US 08/008,001
; FILING DATE: 21-JAN-1993
; APPLICATION NUMBER: US 07/728,783
; FILING DATE: 03-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5686412and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/31853

```

```

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/089,997
; FILING DATE: 06-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 8646
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-742-4343
; TELEFAX: 516-742-4366
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1250 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-441-139-9

Query Match 4.2%; Score 81.5; DB 1; Length 1250;
Best Local Similarity 21.9%; Pred. No. 9.7;
Matches 68; Conservative 34; Mismatches 115; Indels 93; Gaps 14;

Qy 31 EKFSASEIGP---EAVTAFLGIFTTINEHIENAPTGRHVG-----SSGNDPDSLVSPL 81
Db 180 KRISGGEAGVTYKIDVSQADSGNSITEQSDKFPFPPNAHOGHRRATSNLSPFSEKFP- 238
Qy 82 LIAILKDLTVVVAEAEHFYGDKNKNIITLTKAVIRLALFRNSGYKMLQGETPNE 141
Db 239 -----PNSHGDNDEFI-----ATSSHRRSKTRNEY-----SPGINSNW 274
Qy 142 EKSNQSESO-----NRAGNSGRNLGPHGLGNQNHNPWNLEGRAMSALSSFGON 191
Db 275 RNQSQQPQQLSPFRHGRSNDYNSFTLEPPAIFQQGHK--RASNSVHSFSSQGN 332
Qy 192 ARTTSTPCGWSRIQH---OQAVTEPPMIKERRRTMSSELLTEK-----GVNGALF 239
Db 333 -----NGGKRKSLFAPLPQANI-PELIQGRVLVAGILRVNKNRSDAWSTDGALD 383
Qy 240 AIGEVLVITRPLIVLFIRKYGVRMIPWALSIVDTLGLNLSKMWGEKSKQVHFG 299
Db 384 AD-----IYICGSKDRN-----RALEGDLVAVELLEIVDDVWESKR----- 418
Qy 300 PEKDELRRRK 309
Db 419 -EKEEKRRK 427

RESULT 12
US-08-447-500-24
; Sequence 24, Application US/08447500
; Patent No. 5627064
; GENERAL INFORMATION:
; APPLICANT: Hoekstra, Merl F.
; TITLE OF INVENTION: PROTEIN KINASES
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Spensley Horn Jubas & Lubitz
; STREET: 1880 Century Park East, Suite 500
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90067
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/447,500
; FILING DATE:

```

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 524 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-454-097-24

Query Match 4.2%; Score 80.5; DB 1; Length 524;  
Best Local Similarity 22.3%; Pred. No. 3.2;  
Matches 33; Conservative 17; Mismatches 61; Indels 37; Gaps 4;  
QY 74 DPSLSYPLLIALLKDLTVVEVAEAEHFY-----GDKKWNYYIILTEAMKAVIRLALFRNS 127  
DB 305 DPDYDY--LISLMDALRLNDLKDDGHYDMDLNGGKGNKIKINRRAN----- 350  
QY 128 GYKMLQGETPNEEKDSNQSESONRAGNSGRNLGPHGLGNQNHNPWNLEGRAMSALSS 187  
DB 351 -----LHGYGNPNPRVNGTARNNTNSKTRNTPVATPKQAQNSYNKD----- 396  
QY 188 FGQWARTTTSPTGWSRRIOHQQAVIEP 215  
DB 397 ---NSKRISSNFOSFTKQOHLVKKIEP 421

RESULT 14  
US-08-453-866-24  
; Sequence 24, Application US/08453866  
; Patent No. 5756289  
; GENERAL INFORMATION:  
; APPLICANT: Hoechst, Merl F.  
; TITLE OF INVENTION: PROTEIN KINASES  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Spensley Horn Jubas & Lubitz  
; STREET: 1880 Century Park East, Suite 500  
; CITY: Los Angeles  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 90067  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/453,866  
; FILING DATE: 30-MAY-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/008,001  
; FILING DATE: 20-JAN-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Wetherell Ph.D., John R.  
; REGISTRATION NUMBER: 31,678  
; REFERENCE/DOCKET NUMBER: PD-2458  
; TELEPHONE: (619) 455-5100  
; TELEFAX: (619) 455-5110  
; INFORMATION FOR SEQ ID NO: 24:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 524 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-453-866-24

Query Match 4.2%; Score 80.5; DB 1; Length 524;  
Best Local Similarity 22.3%; Pred. No. 3.2;  
Matches 33; Conservative 17; Mismatches 61; Indels 37; Gaps 4;  
QY 74 DPSLSYPLLIALLKDLTVVEVAEAEHFY-----GDKKWNYYIILTEAMKAVIRLALFRNS 127  
DB 305 DPDYDY--LISLMDALRLNDLKDDGHYDMDLNGGKGNKIKINRRAN----- 350  
QY 128 GYKMLQGETPNEEKDSNQSESONRAGNSGRNLGPHGLGNQNHNPWNLEGRAMSALSS 187  
DB 351 -----LHGYGNPNPRVNGTARNNTNSKTRNTPVATPKQAQNSYNKD----- 396  
QY 188 FGQWARTTTSPTGWSRRIOHQQAVIEP 215  
DB 397 ---NSKRISSNFOSFTKQOHLVKKIEP 421

RESULT 15  
US-08-185-359-24  
; Sequence 24, Application US/08185359  
; Patent No. 6060296  
; GENERAL INFORMATION:  
; APPLICANT: Hoechst, Merl F.  
; TITLE OF INVENTION: Protein Kinases  
; NUMBER OF SEQUENCES: 57  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Sears tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/185,359  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/008,001  
; FILING DATE: 21-JAN-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/728,783  
; FILING DATE: 03-JUL-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: No. 6060296and, Greta E.  
; REGISTRATION NUMBER: 35,302  
; REFERENCE/DOCKET NUMBER: 27866/31853  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 24:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 524 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-185-359-24

Query Match 4.2%; Score 80.5; DB 3; Length 524;  
Best Local Similarity 22.3%; Pred. No. 3.2;  
Matches 33; Conservative 17; Mismatches 61; Indels 37; Gaps 4;  
QY 74 DPSLSYPLLIALLKDLTVVEVAEAEHFY-----GDKKWNYYIILTEAMKAVIRLALFRNS 127  
DB 305 DPDYDY--LISLMDALRLNDLKDDGHYDMDLNGGKGNKIKINRRAN----- 350  
QY 128 GYKMLQGETPNEEKDSNQSESONRAGNSGRNLGPHGLGNQNHNPWNLEGRAMSALSS 187

Db 351 -----LHGYGNPRVNGNTARNVNTNSKTRNTTPVATPKQQAQNSYNKD----- 396  
QY 188 FQONARTTTSTPGWSRRRIHQHQAQVIEP 215  
Db 397 ---NSKSRISSNPQSFTKQOHVLLKKIEP 421

Search completed: July 11, 2001, 11:36:34  
Job time: 93 sec

